TAX REVENUE FORECASTING

IN A

DEVELOPING ECONOMY

WITH

SPECIAL REFERENCE TO INDIA

by

DINESH KUMAR SRIVASTAVA

Dissertation submitted to the University of St. Andrews in partial fulfilment of the regulations for the degree of Master of Letters in Economic Development, November 1974.

TAX REVENUE FORECASTING IN A DEVELOPING ECONOMY WITH SPECIAL REFERENCE TO INDIA - ABSTRACT

This study attempts to provide a theoretical framework for the construction and evaluation of tax-revenue forecasts in a developing economy. In addition, a partial empirical counterpart is provided in the study of three major Union taxes in India.

The theoretical framework provides for building tax-interaction models in economies with multiple tax-rates for various direct and indirect taxes. It is noted that revenue estimation in developing economies faces a major problem regarding the way in which discretionary changes in tax-rates and bases should be introduced in the prediction model. Earlier literature in this field suggests two ways of dealing with this problem. One alternative is to construct adjusted revenue series such that the effects of discretionary changes may have been removed from them. The limitations of the theoretical assumptions underlying the available adjustment methods are analysed. The second alternative is to use tax-rates in the regression equations in addition to other exogenous variables. This is theoretically more appealing but has a limited application for taxes with multiple tax-rates for different categories of tax-bases. It may not be possible to use all the relevant tax-rates in the estimation equations because of the implied loss of degrees of freedom for samples of limited size, and because of possible problems of multicollinearity.

In view of these problems it is suggested that tax-rate functions should be estimated so that the rate-structure of a tax can be represented by a limited number of parameters. These parameters can later be used in revenue-estimation equations. Using tax-rate functions, a model specifically allowing for interaction among tax-revenues is developed. It is recognised that the complexities and special characteristics of individual case studies necessitate additional modifications to this framework. Some typical problems in the case of developing economies are identified and suggestions made regarding possible methods of dealing with them.

D.K. SRIVASTAVA.

Declaration

I hereby declare that the following thesis is based on research carried out by me, that the thesis is my own composition and that it has not previously been presented for a Higher Degree.

The research was carried out in the University of St. Andrews after my admission as a research student and M.Litt. candidate in October 1972.

D.K. SRIVASTAVA
(Candidate)

Certificate

I certify that the aforesaid candidate has fulfilled the conditions of the Ordinances and Regulations prescribed for the degree of M.Litt.

Dr. G.K. SHAW
(Supervisor)

TAX REVENUE FORECASTING IN A DEVELOPING ECONOMY WITH SPECIAL REFERENCE TO INDIA

Acknowledgements

This work has matured under the supervision of Dr. G.K. Shaw. He originally suggested the subject and saw it to fruition by variously guiding, criticising, and correcting me at all stages of the work. To him I am deeply indebted.

Very special thanks are due to John Hey, who read through the entire manuscript and made valuable suggestions for its improvement. He helped me negotiate some of the slippery edges of an econometric exercise of this nature.

I thank Prof. P. Robson for continually encouraging me with the work. Thanks are also due to S. McDowall, R.S. May and W.V. Thomas who stimulated my interest in the study of the problems of developing economies.

During the course of this work, I had occasion to discuss various aspects of the subject of tax-revenue forecasting with Prof. W. Vickrey of the United Nations, Prof. L. Harris of the Tax Foundation Inc., and Prof. Richard Goode of the International Monetary Fund. They freely provided me with thoughts and materials on the subject. I also thank Sheetal Chand of the I.M.F. and Ved. P. Gandhi and Mrs. Kazuko K. Artus of the World Bank for devoting considerable amounts of time to discussing my work and making valuable suggestions. Discussions with Malcolm Gillis at the Harvard Institute of International Development were singularly stimulating.

Data and other relevant research material were primarily obtained from the Library of the Indian High Commission in London and the Joint Bank-Fund library in Washington, D.C. I acknowledge the useful assistance and facilities provided by these bodies.

This work was completed during a period in which I held a Commonwealth scholarship in the United Kingdom. I thank Commonwealth Awards Commission in the U.K. and the British Council for respectively awarding and administering the scholarship.

Completion of this work and continuing research in the field of forecasting in developing economies has been made possible by a grant from the Social Science Research Council of the United Kingdom. I acknowledge this financial assistance.

I thank Mrs. Ellen Graves for typing a systematic work out of what looked at first sight to be a haphazard mess of tables and equations.

In the end, I thank Neil Gilchrist who laboriously went through the typescript and saved me from commiting many errors. I also thank Suresh T. Gore who checked the tables and equations.

Without the generous help and assistance of all these people, this work would not have been possible. For all the shortcomings and mistakes that remain, I am alone to blame.

CONTENTS

List of Tables:

CHAPTER 1

INTRODUCTION

1.1 Forecasting, Taxes, and Developing Countries

Forecasting represents the application of a set of techniques for predicting future, or more generally, 'beyond-sample' values of given variables. Tax-revenues represent a set of such objective variables and developing countries represent a context. Putting these three things together, i.e. the application of forecasting techniques for predicting future tax-revenues in a developing economy, would broadly indicate the concern of the present study. However, forecasting techniques have evolved, by and large, in studies related to advanced economies, and hence one begins by asking whether they can be simply wheeled into a new territory or whether the turns and twists peculiar to developing countries need to be carefully negotiated.

The answer to this lies in making a distinction in the application of prediction techniques between what may be called its 'mechanical' and 'judgemental' components. To a limited extent some techniques have a purely mechanical content in that having fed in some data, from whatever context they may be derived, certain results would mechanically and automatically follow. Thus, for example, to a large extent, techniques for evaluation of forecasts need data on two series: predictions and their corresponding realizations. Given these, the same set of techniques can be applied and the same type of conclusions can be derived, irrespective of whether the context is a developed or a developing economy. Similarly, a number of extrapolation techniques are purely mechanical. But this is the less demanding half of the story. For the most part forecasting is an exercise in judgement in that the forecaster has to exert a choice at various stages of the endeavour. For example, he has to choose from among alternative prediction techniques; given a technique, from among alternative possible models and explanatory variables; given a model, from among possible estimation techniques and so on. It is at these steps that careful choice needs to be exerted so as to accommodate the special characteristics of a developing economy.

1.2 Utility of Revenue Forecasts

There are at least three parties who will have a more than academic interest in the future resource position of a developing economy vis-a-vis its tax-revenues, viz., the government in a developing economy which is actively engaged in formulating development plans; external and international aid agencies which have to formulate their aid-programmes considerably in advance; and governments in advanced economies which would like to reflect on their aid-programmes and trade prospects in the light of the tax-prospects of their client economies in the developing world.

The major interest in tax-revenue forecasts lies, of course, with the planning agency of the country concerned. Government expenditures in a developing economy are conditional upon the past, present and anticipated performance of tax-revenues. To effectively plan for economic development, the governments would be interested not only in the current tax prospects of the country but also in its future tax prospects.

Apart from their relevance for general economic planning, revenue forecasts will be of more direct interest to fiscal planning authorities in view of the fact that such forecasts can be used for devising fiscal policy and tax-reforms. This would be possible if conditional forecasts can be made of the effect of alternative tax policies on tax yields.

In this context, one distinguishes between two components of revenue growth in an economy, viz., the automatic component and the discretionary component. The discretionary component accounts for the effect on tax-revenues of specific policy changes brought about by legislative or administrative action. These encompass changes in tax-rates, tax-bases, administrative structure of collection of tax-revenues, and so on. These changes are treated as exogenous. The automatic change in the tax-yield is one that would occur in the absence of discretionary changes. It would, therefore, occur because of the general economic development of the economy. While the revenue-impact of major administrative reforms brought about by legislative action are covered by the discretionary component, growth in tax-yields due to normal improvement in administrative efficiency would be counted in the automatic component.

The task of the forecaster is to predict both kinds of changes. In practice both of these would boil down to simulation exercises. The automatic growth in tax-yield would be predicted on the assumption that a given legal system and tax-rate structure for a tax would prevail in the period for which the forecasts are being generated. The discretionary growth would be predicted on the assumption that certain policy configurations are introduced in the economy. The actual experience of the economy would, in fact, be a combination of the two situations. However, the simulation exercises based on their respective hypothetical detours would still be worthwhile as long as it can be hoped that the actual experience would be shaped, at least in part, by these conditional predictions and that the planners, both fiscal and economic, would have been worse off in their capacity as decision-makers in the absence of these predictions.

Apart from automatic and discretionary growth in tax-yields, forecasts regarding tax-shares, i.e. the proportions of individual taxes to total tax-revenue in any given year, will also be of direct relevance to tax-planning authorities inasmuch as tax-shares shed light on the tax-structures of an economy. Until now academic interest has centered in conducting cross-section studies of tax-structures via a study of tax-shares. Comparisons have been made, for example, of the share of total direct taxes, total indirect taxes, personal income taxes, etc. between countries at different stages of development thus indicating the nature of change in the tax-structure that would go with economic development. Projections of tax-shares over time in a given country can be compared with the present tax-structures in advanced economies to draw some useful conclusions.

Tax-share forecasts may be derived either from projections of individual tax-revenues or directly from historical movements in tax-shares. Projections based on historical movements either in revenues or tax-shares would mirror the state we should expect in the absence of active interference from tax-planning authorities. If the future image of the tax-structure does not correspond well with the existing tax-structures of developed countries, then some idea of the direction of required changes, allowing for the special circumstances of individual countries, would be obtained from tax-share forecasts.

1.3 Short-, Intermediate- and Long-term Forecasts

The current and universally encountered revenue forecasting exercise in a developing economy remains the traditional budget forecast which relates to one year or part thereof. These forecasts are typically based on the opinions of the fiscal experts of a country's treasury or ministry of finance or some such body generally with no explicit or formal forecasting model behind them.

This is not to say that such forecasts will not generally be found sufficiently accurate except in cases where over- or under-estimates may be intended for political reasons. These forecasts may also at times not distinguish between automatic and discretionary components of revenue growth and provide tax projections based on the package of past tax policies and changes proposed in the current year.

While such short-term forecasts remain important, they cannot serve economic planning purposes. The planning exercise is based on a development plan, the average length of which is four to five years. For this purpose we need intermediate-term forecasts as also a greater emphasis on the automatic growth of tax-revenues, i.e. changes in tax-revenues that may be due to the contribution of planned economic development itself.

Intermediate-term forecasts would abstract from short-term cyclical processes that tend to cancel out over a longer period. Since, in the medium term, the forces affecting tax-revenue and economic growth would generally be more streamlined than in the short-term, it is possible to rely on explicit, clear-cut forecasting models. Important as the medium-term forecasts may be, such forecasting exercises are seriously lacking in developing countries.

Ultimately long-term projections would also be of considerable importance. Long-term changes in tax-revenues would depend on structural changes in the economy which may be very difficult to predict. Although a good economic plan should make adjustments for economic processes that generate their effects in a period bigger than the normal planning horizon, predictions regarding these are difficult and work in this direction is not very advanced even in developed economies.

1.4 Alternative Techniques of Forecasting

Analytically, various techniques of forecasting can be grouped into three broad categories, viz., techniques that exploit relationships between two or more variables in a one-way causation; techniques that exploit relationships between two or more variables in a multi-way causation; and techniques that are based on the historical behaviour of one variable alone and do not exploit any economic relationships. These techniques can respectively be called partial equilibrium forecasting, general equilibrium forecasting and pure extrapolation.

The postulate behind partial equilibrium forecasting is that the variable to be predicted, called the dependent variable, is affected by a number of variables, called independent variables but that the independent variables are not, in turn, affected by the dependent variable. The preliminary choice of independent variables would be governed by the knowledge of their economic relationships with the dependent variable, and the final choice would be subject to the statistical strength of these relationships when tested over a historical sample period. One difficulty with partial equilibrium forecasting is that it does not produce consistent results. For example, the sum of individually predicted tax revenues would not equal the independently predicted total tax revenue.

This is also a difficulty with pure extrapolation: in this, we study, over time, the behaviour of the variable to be predicted and use the knowledge only of the past values of this variable to predict its future values.

General equilibrium forecasting does away with the problem of consistency by introducing a number of identities and by relating variables to each other by a set of equations. Here, the whole macroeconomy is represented by an equation system. For practical reasons, even this model would not be truly general in that it will still have a number of variables, called <u>exogenous</u> variables, which will be determined outside the model. The model-determined variables, called <u>endogenous</u> variables can and normally will be related in a multi-way causation framework. In addition, in any estimating equation, a number of lagged endogenous variables can be introduced wherever desirable. Forecasts would then be made from the <u>reduced-form</u> of equations -- which means solving the equation system for endogenous variables such that each endogenous variable is represented as a function of exogenous and lagged endogenous variables which together are named as <u>predetermined variables</u>. It will be possible to generate forecasts by using externally predicted future values of exogenous variables as in the case of partial equilibrium models. The values of lagged endogenous variables will be generated by the model itself for periods beyond the sample.

The choice between these alternative techniques is a difficult task. In the end, this choice would depend on the costs of employing a particular technique and on its forecasting performance as compared to that of other techniques. This performance is judged, in part, by the deviation of prediction from realization and several measures for this purpose have been proposed in the vast literature on this subject. However, a choice on this ground can only be made when data about realizations is available.

On *a priori* grounds, general equilibrium forecasting would seem to be the best, if costs were not an important consideration. This is so because it uses maximum information in producing a forecast and is based on the knowledge of economic relationships. If variables in an economy are interrelated and if economic theory provides correct insight into the nature of these relationships, forecasts that are made on this basis should be better than those that are based on purely mechanistic grounds as in extrapolation,or on incomplete relationships as in partial equilibrium models.

In practice, however, general equilibrium models do not produce results that are consistently better than those of the simpler techniques of extrapolation and partial equilibrium models. Much of this may be due to specification errors, errors in measurement of data, etc., and analytically the differences in forecasting performance cannot be assigned to the 'pure' merits or demerits of alternative techniques.

As a matter of fact, comparison between techniques may not be possible at all. This is because under each technique, a family of models can be formulated. All that is possible is to compare the performance of one specific model of one family to another specific model of another family. Each model is the resultant of an interaction between sample data, known economic relationships, statistical tests and the forecaster's commonsense. Comparison of specific models cannot, therefore, be taken to reflect, in a rigorous sense, the merits of the family to which they belong.

In addition, different forecasting techniques should not be taken as mutually exclusive. Techniques have been proposed (Bates and Granger, 1969; Nelson, 1972) by which forecasts from different methods can be combined such that the resultant may perform better than the individual forecasts. This is so because the predictive power of historical information used in different forecasts can be bettered by combining them whenever individual forecasts have not been able to fully exploit the predictive power of the data they have used.

1.5 The Scope of the Present Work

The scope of the present work is limited to partial equilibrium forecasting. The considerations in the previous sections have been introduced with a view to providing a perspective to this. We have noted that it is desirable to experiment with general equilibrium econometric models and derive from them, as a part, revenue-forecasts so that interaction among taxes and interaction of taxes with other macro-variables of the economy may be allowed for. But this is a very broad study and outside the scope of this work. In general large econometric models may yet be unsuitable for developing economies because of high costs and inadequate data, wherever this may be the case.

Within the partial equilibrium setting, it is intended to develop a framework from which conditional forecasts of tax-revenues can be generated for alternative values of tax-rate parameters. This framework should also be able to predict the automatic growth in revenues when tax-rate parameters are held constant at the level for some given year. The appropriate forecasting period seems to be a medium term extending to four to five years such that it coincides with the average length of a development plan.

In addition, it is intended to study techniques of evaluation of forecasts and apply them to evaluate annual budget forecasts regarding taxes. These are one-year-ahead forecasts prepared by the ministry of finance or some such body in most developing economies under legal obligation. The basis of these forecasts is generally an informal and implicit model aided by expert opinions. Their evaluation will normally point to certain systematic errors and once the forecasting agency becomes aware of these, it can search for possible sources of error. Systematic errors are easily removed and it can be contended that budget forecasts in most economies can be considerably improved if they are properly evaluated after data about actual tax-revenues becomes available.

1.6 The Premises of Partial Equilibrium Forecasting of Taxes

In the partial equilibrium framework revenue-estimates of individual taxes are visualised as functions of a number of predictor or explanatory variables. The relationship between the predicted variables and the predictor is estimated with the help of data related to these in a given sample period. Given the estimated relationships, one has to feed in independently predicted values of exogenous variables for the forecasting period to generate forecasts for the endogenous variables, in this case, the individual tax-revenues.

The assumption behind the partial equilibrium premise is that the nature of causation between growth in exogenous variables (say, income) and growth in endogenous variable (say, sales-tax revenue) is <u>one-way</u>, i.e. growth in income affects growth in tax-revenue but not vice-versa. Further, projection of individual tax-revenues separately does not allow a framework in which yields of different taxes may be seen as interacting.

These are drastic assumptions but the alternative is a more general equilibrium framework, the feasibility and desirability of which, as already noted, is limited in the context of developing countries. However, it is possible to extend the analysis so as to allow interaction among taxes while still treating other important macro-variables such as income, sales of commodities, imports, exports, etc., as determined outside the model. A truly general equilibrium model should treat these variables as endogenous. Thus, allowance for interaction among taxes makes the model less restrictive but it still remains a highly partial approach. For the moment, however, we concentrate on problems of forecasting individual taxes separately. In each case we have to determine the choice of predictor variables.

1.7 The Mechanics of Revenue Generation

For taxes which are levied on commodities (e.g. sales tax, excise tax, import tax), the formulation of a revenue-generation equation is simple. Consider, for example, a sales tax. Let there be J tax-rates each pertaining to a distinct type of consumption expenditure. Let total taxable sales in the jth class, i.e. after exemptions, etc., are deducted, be C_j and the rate pertaining to this class be r_j. Then the sales tax-revenue (R_s) can be written as

$$(1.1) \qquad R_s = \sum_{j=1}^{J} r_j C_j$$

Similar equations can be written for other taxes on commodities.

Taxes on income (e.g., personal income tax, corporate income tax) can also be written in this form. But for a clearer insight one can start from an earlier step.

Suppose there are J income-tax rates (r_j) which become operative at defined income-slabs such that for an individual in the highest income-bracket total tax is given by

$$(y_1-y_0)r_1 + (y_2-y_1)r_2 + \dots \dots + (y_i-y_J)r_J$$

where y_0 is the level at which incomes become taxable; y_1, y_2, ...y_J define levels at which a new rate becomes operative; and, y_i is the taxable income of the individual. Taxes for individuals in lower brackets can be written in a similar fashion such that the total income-tax will be given by the sum of the terms in the following:

For an individual with taxable income	Tax
$\geq y_J$	$(y_1-y_0)r_1 + (y_2-y_1)r_2 + \ldots\ldots\ldots\ldots + (y_i-y_J)r_J$
$\geq y_{J-1}$	$(y_1-y_0)r_1 + (y_2-y_1)r_2 + \ldots\ldots+(y_i-y_{J-1})r_{J-1}$
.	
.	
.	
.	
$\geq y_1$	$(y_1-y_0)r_1 + (y_i-y_1)r_2$
$\geq y_0$	$(y_i-y_0)r_1$

Summing up vertically, total personal income tax can be written as

(1.2) $$R_y = \sum_{j=1}^{J} r_j Y_j$$

where Y_j is interpreted as total personal income taxed at the rate r_j.

Thus, both taxes on income as also those on the value of commodities can be written in the same form. It should be noted that in the above formulations we abstract from exemptions that are allowed after tax is assessed.

The expressions for tax-revenues are simpler when only a single tax-rate prevails. For example, if there were just one sales-tax rate (say, r_s) for all types of consumption expenditure we would write the sales tax-revenue as

(1.3) $$R_s = r_s C_s$$

where C_s is total taxable consumption expenditure.

In the case of a single tax-rate the tax-revenue would depend on the total tax-base, i.e. C_s in the case of sales-tax apart from the tax-rate itself. In the case of multiple tax-rates for a given tax, the revenue would depend, apart from the vector of tax-rates, upon not only the total tax-base but also its distribution among various rate-classes.

Whether we have a single or multiple rates, an important part of the forecasting exercise is to predict growth in tax-bases. This necessitates a search for exogenous variables that explain growth in tax-bases. We may start with some general considerations.

1.8 The Choice of Predictors: General Considerations

Traditionally, a vital characteristic of a tax-system is considered to be the response of tax-revenues to changes in income. This response is measured by concepts such as income-elasticity of a tax, income-buoyancy of a tax and marginal tax-income rates. These measures provide an indication of automatic growth in tax-revenues since they relate growth in tax-yields to economic development as indexed by some concept of income (G.N.P., per capita income, etc.). Consequently, comparisons of tax-elasticities and allied concepts form a basis for comparison of different tax-systems and one frequently comes across studies of this nature. A next step forward has been to exploit the tax-income relationship for forecasting purposes. Some concept of income which stands as an index for economic development seems therefore a major explanatory variable to be considered.

The influence of income on tax-yields should be visualised through its effect on tax-bases. The nature of this effect can be considered in specific cases. The base for personal income tax, for example, would increase with economic development because people would move from (i) non-taxed to taxed income-brackets, and from (ii) low-taxed to high-taxed income brackets. The same will be true in the case of corporation income-tax.

The base for sales taxes would go up with increases in consumption expenditure following the increase in incomes. The nature of the relationship between the sales tax-base and income is expected to be positive; however, the precise relationship would depend on people's marginal propensities to consume taxed and non-taxed items as also high-taxed and low-taxed items. The same should be true for excise taxes.

The effect of increased incomes on revenue from import taxes will be the resultant of a mix of a positive and a negative effect. As incomes increase, demand for imports are expected to rise dependent on the marginal propensity to import. On the other hand, as economic development takes place, developing countries undergo bouts of import substitution, thus undermining the base of import taxes. For this reason again, we will expect consumption expenditure on domestic goods to increase at the cost of imported goods and thus expect the bases of sales-tax and excise taxes to increase.

The base of export tax seems less dependent on growth in domestic income. Apart from internal supply factors, exports of a country depend on world demand and price factors. Income would serve as an appropriate explanatory variable only to the extent it affects internal supply factors, and these, in turn, affect growth in exports.

For reasons cited above the nature of the relationship between income and tax-yields would differ from tax to tax. The performance[1] of income-based forecasts of tax-yields would depend on the significance of the estimated relationship and its explanatory power.

In most cases other exogenous variables will have to be introduced in addition to income. For example, population may be introduced as an explanatory variable in the personal income-tax equation to take account of the increase in the number of tax-assesses due to an increase in population. It may also explain some part of increase in expenditures on commodities and thus may be introduced in sales and excise tax equations. Total imports may need to be disaggregated into capital and consumption goods in the import tax equation and so on. In each case the choice of explanatory variables needs detailed consideration and the appropriate stage to deal with these is when individual taxes are taken up.

---

1. Some idea of the nature and strength of tax-income relationships may be obtained from the cross-section studies of Musgrave (1969) and Due (1970). With a cross-section of countries at various stages of economic development the general idea was to deduce some conclusions about the nature of growth in taxes as economic development takes place. The independent variables, however, are tax-shares rather than tax-yields and the equations cited below are purely for illustrative purposes. These equations are from the Musgrave study. Similar equations can be obtained from the Due study.

Dependent Var.	Ind. Var.	Constant	Regression Coeff.	R^2
T_p/T	Y_c	.1981 (7.574)	+.000137 (5.622)	.47
T_{cp}/T	$Y_c > \$600$	.01318	+.0000959 (3.12738)	.48
$\log (T_{id}/T)$	$\log Y_c$	4.986 (25.269)	-.1761 (-5.348)	.43
$\log (T_{cd}/T)$	$\log Y_c$	1.79 (7.0312)	-.663 (-4.425)	.39

Key: T_p - personal income tax; T_{cp} - corporation tax; T_{id} - indirect taxes; T_{cd} - customs duties; T - total tax-revenue; Y_c - per capita income; $Y_c > \$600$ - the sample consists of countries with per capita income greater than \$600. Figures in brackets are t-ratios.

No matter how many exogenous variables are introduced, the procedure remains one of estimating historical relationships by fitting regression equations. The estimation of regressions coefficients and the fact that the choice of explanatory variables is aided by statistical tests after preliminary choice has been made on theoretical grounds, leads one to anticipate some of the difficulties that might be faced in applying standard statistical techniques in the context of developing countries. It is appropriate to recognize these problems at the outset so as to guard against avoidable mistakes in practice.

1.9 Data Problems

Prediction of tax revenues requires time-series of tax-revenues, extensively disaggregated data for tax-bases and for factors affecting tax-bases. It is difficult to imagine that the data-base provided by developing countries can satisfactorily meet all these requirements. Data-collection in developing countries is still a very haphazard exercise. In many cases it is not geared towards specific objectives but is merely a mechanical exercise for publishing certain series that might currently be published in developed countries. Even where a long-run series is available, one is suspicious of the ad-hoc adjustments and estimates that might have gone into it at one time or another for filling in the blanks for components not available at the right time. Many of these adjustments lie screened behind a neat-looking final series and are not made explicit.

An inadequate data-base renders any econometric exercise precarious. For our purposes, however, certain limitations can be noted.

First, data are not consistent and uniform over time. This makes the task of finding a respectable sample period considerably difficult. For some years, one may find that a vital series is punctured right inside the stipulated sample; somewhere within the sample years one or two series may have been revised; disaggregated data may stop considerably short of other series, and so on. Many of these problems can, of course, be glossed over by making appropriate and careful adjustments, but the more one does so, the less reliable the results become.

Second, even where adjustments are made, the sample-periods for the estimation of equations in the model may still be very limited. For some developing countries, regression equations have been estimated for as small a sample period as five to seven years.[1] The inadequacy of the length of time-series raises doubts as to the reliability of estimated coefficients and renders the use of sophisticated statistical techniques virtually useless. It also limits the number of exogenous variables that can be introduced in an estimating equation since the degrees of freedom in the estimation depend on the size of the sample, i.e. number of observations, and the number of regressors. The problem of a small sample size, however, recedes in significance with every year that goes by as it adds, under normal circumstances, one more year to an existing sample.

Finally, in developing countries, models become 'dated' very quickly. This is because national accounts in these economies are constantly undergoing revisions. The revisions not only affect individual coefficients but also the overall conclusions that derive from a model.

When there is a complete changeover in the technique of

1. See, for example, UNCTAD (1968) and ECAFE (1968)

estimation of national income, a series based on one technique would, of course, be not comparable with another unless estimates of past years are recalculated with reference to the new method. Even when there is not a complete change of techniques, the national income estimates of current income in developing countries improve year by year with general improvements in administrative efficiency and wider coverage, etc. Where, as a result of this, current data improve in the sense of having a more comprehensive coverage of the economy, the estimates of income in the past year become, in fact, understatements of income in comparison to the current estimates. This obviously affects the estimates of the income-response of revenues.

These problems are formidable. But it should be noted that they are not applicable in a blanket way to all developing countries. One needs to distinguish between developing countries where national accounts are in an advanced stage of sophistication and those where they are not. Furthermore, exercises in model-building in developing countries are useful, if not entirely in their results, at least in pointing out the directions in which the data-base should develop for enabling one to get more reliable results.

1.10 Methodological Problems

For illustrating some of the methodological problems in a forecasting exercise such as the present one, let us use a simple multiple regression model. Suppose we want to predict a variable y using two predictors x_1 and x_2. Assume, for the moment, that a linear relationship exists between the predicted variable and the predictors such that the 'true' relationship can be written as

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t \tag{1.4}$$

where β_0 is a constant, β_1 and β_2 are 'true' partial regression coefficients, and u_t is a random disturbance term independent of x_1 and x_2. Suppose the distrubance term is normal with zero mean and given variance.

Suppose, further, that a sample of T annual observations for y, x_1 and x_2 are available such that the constant term and the regression coefficients can be estimated. Suppose the least-squares estimates for these are b_o, b_1 and b_2.

We can now generate a prediction for y_{T+1} by

$$(1.5) \qquad y_{T+1} = b_0 + b_1x_{1,T+1} + b_2x_{2,T+1}$$

Here we have replaced u_{t+1} by its expected value which is zero and used sample-estimates of β_0, β_1 and β_2. The prediction will, of course, be possible only if predictions regarding x_1 and x_2 for the period T+1 are available.

A number of problems can now be noted in a forecasting procedure such as above.

(i) Specification Errors

The first difficulty arises from the possibility of incorrect specification of the regression equation. The 'true' system may contain more explanatory variables than x_1 and x_2; or, one of these may be irrelevant; or, the form of the relationship may not be linear. These errors occur because of insufficient knowledge as to the system which generates the objective variable. Many a time deficient data may itself lead to the use of inappropriate relationships.

Specification errors frequently occur when macroregression models are used and they have significant implications for the analysis. The implications vary with the type of error.

If relevant explanatory variables are omitted, when they should have been included, the least squares estimates for the included variables become biased[1] if they are correlated with variables that have not been included. This type of error is less important when the purpose of analysis is pure forecasting than when the analysis also has policy implications. Suppose a correct specification of equation (1.3) requires the inclusion of a variable x_3 which is correlated with x_1 and x_2. If the equation is estimated only with x_1 and x_2, the estimates will be biased. If forecasting alone is the purpose, one may still work with the misspecified relationship as long as one has reason to believe that the relationship between x_1, x_2 and x_3 would remain the same in the forecasting period as in the sample period. Under this assumption one may let x_1 and x_2 do the job for x_3 as well. But when the analysis has policy implications, it may be too risky to rely upon the biased coefficients.

Similarly, the functional form of the equation needs to be carefully chosen. Each form has definite economic assumptions behind it, and it is not a matter of indifference as to what form is chosen. For example, a linear relationship with an intercept, such as (1.3), implies that the marginal rates ($\partial y/\partial x_1 = \beta_1$; $\partial y/\partial x_2 = \beta_2$) are constant; a log-linear relationship such as

$$(1.6) \qquad y = \beta_0 x_1^{\beta_1} x_2^{\beta_2}$$

implies that the partial elasticities of y w.r.t. x_1 and x_2 remain unchanged. Hence, the chosen functional form should bear with the hypothesized economic relationship.

For this reason, one should not rely totally on statistical tests and consider that form the most appropriate which gives the

1. See, for example, Johnston (1963, 1972)

'best fit'. If the 'statistical' superiority of one form over another is marginal, there is no reason to choose, in every case, that form which happens to be slightly superior in the particular sample being used because there is no reason to believe that the superiority would persist if the sample was changed. In particular, the economic implications of the form and the objectives of the analysis should be borne in mind before making a choice.

(ii) Multicollinearity

Standard multiple regression analysis tells us that in an equation of the type (1.4) an unambiguous meaning to the estimated coefficients b_1 and b_2 can only be attached, assuming correct specification and reliable data, when the explanatory variables x_1 and x_2 are causally independent or 'orthogonal'. On the other extreme, if x_1 and x_2 are perfectly 'collinear', regression coefficients cannot be estimated.

Normally, the situation encountered in a developing economy is one where there is a high degree of collinearity among the regressors. This is so because of the presence of strong trend components, the use of highly aggregated variables and limited sample periods. The presence of multicollinearity among regressors makes the coefficients extremely sensitive to sample coverage, data errors, etc.

The argument in the case of two regressors (say) can be put forward as follows. The least square estimate of the partial regression coefficient b_1 can be written as[1]

$$b_1 = \frac{b_{y1} - b_{y2}b_{21}}{1 - r_{12}^2} \tag{1.7}$$

1. See, e.g., Walters (1968, 1970). For formulation of the multicollinearity problem in terms of its effect on the sampling variance of estimated coefficients, see Johnston (1963, 1972).

where b_{y1} is the regression coefficient in the simple regression of y on x_1; b_{y2}, that in the regression of y on x_2; and b_{21}, that in the regression of x_2 on x_1. r_{12}^2 is the coefficient of correlation between x_1 and x_2.

Since $1 - r_{12}^2$ occurs in the denominator, the coefficient becomes highly sensitive to changes in r_{12}^2 when they are near unity. For example, a change in r_{12}^2 from 0.98 to 0.99, which may occur just because of sampling and rounding errors, will have the effect of doubling the estimate of b_2, assuming the numerator is not affected. Thus, in the case of highly collineal regressors, i.e. regressors with r_{12}^2 near unity, it is very difficult to rely on the estimated coefficients.

(iii) <u>Reliance on Statistical Tests</u>

The three main statistical tests that are designed to highlight the inadequacies of the model are the t-ratios, the coefficient of determination (R^2), and the Durbin-Watson statistic. The t-ratios point out the statistical significance of the estimated coefficients; the coefficient of determination helps to outline how far the regressors explain the dependent variable and the Durbin-Watson statistic points out if the residual terms are autocorrelated. If so, one needs to check whether any systematic elements are left out from the analysis. One would hope that with such a battery of measures one is adequately guarded against possible misuse of regression models. But as Shourie (1972) points out the situation is far from 'reassuring', when the tests are 'mechanically' applied in the context of developing economies.

The coefficient of determination is not a good guide to the adequacy of specification if spurious correlations exist between variables. In a developing country, strong trend and cyclical components are present in almost any economic time-series. High correlations between variables may, therefore, occur even when there are no causal relationships. A high value for R^2 would in these cases be a suspect guide to the choice of regressors because there is no reason to expect that the same strong relationships would continue beyond the sample period. Thus, the less evidence there is for the existence of a 'causal' relationship, the less sure one should be that two series will continue to move together simply because they have done so in the past.

The t-statistic is a ratio between the value of a regression coefficient and its standard error of estimate. For testing the statistical significance of a coefficient a critical value of t is chosen depending on the degrees of freedom. In the use of t-statistic, one needs to distinguish between situations where the purpose is to judge the statistical significance of the coefficients and where the purpose is also forecasting substantially ahead in the future. When the purpose is forecasting, a stricter value of this benchmark is needed than otherwise.

The Durbin-Watson statistic is designed to test if there is any strong autocorrelation among the residual terms. The presence of autocorrelation in residuals implies that some systematic element has been left out and hence the search for a new explanatory variable is necessary. However, the difficulty in the application of D-W statistic is that it needs a fairly long sample before competent results can be obtained. This is almost a prohibitive limitation when one is working with annual data in developing countries.

(iv) Exogenous Variables

We notice in equation (1.4) that a forecast for period T+1 can be generated only when independently predicted values of exogenous variables for this period are available. Further, the quality of predictions regarding the exogenous variables streamlines the quality of forecasts being generated. It is said with truth that 'no econometric forecast can be better than its data -- in economic jargon its predetermined variables'[1].

When forecasting in a developing economy, it is easy to run into two problems: first, it may be difficult to obtain forecasts regarding the exogenous variables one might be using; and second, where obtainable, these may be of very doubtful quality.

In most developing countries there would be no forecasting institutions and agencies which systematically provide independent macroeconomic forecasts based on formal and scientific methods. Forecasts made by government agencies may be based on informal analysis and may be biased for obvious political reasons. In comparison, in advanced economies, the situation is more fortunate in that for most of the exogenous variables in a partial equilibrium exercise, a choice would be available from among the forecasts being generated by independent bodies with the help of large econometric models. This is to say that forecasting performance in developing countries can be considerably improved only when this kind of a forecasting 'infrastructure' is available.

---

1. Streissler (1970), p. 19.

CHAPTER 2

ECONOMETRIC MODELS FOR REVENUE FORECASTING

Although ultimately it may be desirable to develop macro-econometric models which allow interactions between variables and derive from them, forecasts of tax-revenues as a part of a bigger forecasting exercise, our approach here is less extensive. In developing economies, where an adequate data-base for large econometric models may still be in the formative stage, this would seem a relevant approach to begin with.

Considerations in this chapter lead to the formulation of a revenue forecasting model within the partial equilibrium frame-work. However, to place this model in its proper perspective we start off with an evaluation of a number of simpler models available in the tax-forecasting literature. Limitations of individual models and the direction of subsequent modifications are pointed out. Finally, certain additional considerations are introduced which may be useful when the suggested model fails to encompass the complexi-ties of an empirical situation.

2.1 Traditional Constant Income-Elasticity Models

The relationship between tax-revenues and income has been consistently exploited for predicting tax revenues. One of the simplest ways to predict tax revenues is to assume that income-elasticity of tax-revenues is constant. Simple predictions based on this assumption can be made from

$$(2.1) \qquad R_{t+1} = R_t \left(1 + e_y \frac{Y_{t+1} - Y_t}{Y_t}\right)$$

where R and Y refer to tax-revenue and income respectively and e_y is the elasticity of R with respect to Y. The subscript refers to the time-period. Predictions regarding income would be needed independently.

Traditionally, income-elasticity of revenue is estimated from the regression equation

$$(2.2) \qquad \log R = \log A + e_y \log Y$$

where A is a constant.

Early attempts to estimate income-elasticity such as Groves and Kahn (1952) and Soltow (1955) were based on an equation like (2.2). A number of difficulties can at once be noted in a forecasting framework such as this.

First, the estimation of income-elasticity by (2.2) assumes that the form of the tax-income relation is log-linear. This need not always be the most appropriate form.

Second, the estimate of income-elasticity in (2.2) can be misspecified because it does not reflect the definition of revenue generation, viz.,

$$(2.3) \qquad R = rX$$

where r is the tax-rate and X is the tax-base.[1] Equation (2.2) would be correctly specified only if both r and X were functionally related to income. But, in fact, both r and X are determined, in part, by factors other than income, especially by discretionary or legislative action. Predictions based on the historical relation between R and Y

---

1. In the case of multiple tax-rates, these are to be interpreted as vectors.

would be valid only on the assumption that the discretionary changes regarding rate and base in the sample period would replicate themselves in the forecasting period. In general, this is not a valid assumption.

One method[1] of getting round this problem came in the substitution of R by a series of adjusted revenue-data (say, R_a) such that it was stripped of the effects of discretionary changes in the tax-rates in the sample period. The adjustment is made such that the revenue series is reduced to a constant rate-structure throughout the entire sample period.

Provided such a series was available, it was possible to think that this would be functionally related to income alone. Hence, income-elasticity was estimated from[2]

$$\log R_a = \log A + e'y \log Y \qquad (2.4)$$

Projections based on an elasticity estimated with the adjusted series would predict 'automatic' growth in revenues conditional upon the assumption that the rate-structure of the year to which revenues in all other years have been adjusted would prevail in the forecasting period.

The difficulty in this procedure arises from the techniques used for data-adjustment. These will be taken up in the next section. For the moment, we may note that the adjustment is generally made under the assumption that any given percentage change in rate and base would generate a proportionate change in tax revenues, i.e. elasticity

---

1. For example, Prest (1962), Berney (1970).

2. To distinguish between the elasticity estimates with adjusted and unadjusted series, sometimes the elasticity with the unadjusted series (e.g. in equation (2.2)) is called the 'buoyancy' of tax-revenues.

of revenue with respect to rate and base is equal to one. The validity of this assumption is highly suspect. It was then suggested that rather than construct an adjusted revenue series, it might be best to use the tax rate (Wilford, 1965) and tax base (Ray, 1966) as independent variables. Predictions would then be made from

$$\log R = \log A + e_y \log Y + e_r \log r + e_x \log X \tag{2.5}$$

The model implied by (2.5), i.e.

$$R = AY^{e_y} \cdot r^{e_r} \cdot X^{e_x} \tag{2.6}$$

can be used to study the implications of the procedure of revenue-adjustment.

The implication behind a constant-rate series is that e_r and e_x are equal to 1. If, in fact, this is so, the adjustment procedure would be justified. For, then we can write

$$\frac{R}{rX} = AY^{e_y} \tag{2.7}$$

In this case, the use of an equation such as (2.4) would be permissible. This equation, however, would be misspecified if e_r and e_x are not equal to 1. In this case, we will have

$$\frac{R}{rX} = A\, y^{e_y} \cdot r^{(e_r-1)}\, X^{(e_x-1)} \tag{2.8}$$

Thus, as Berney and Frerichs (1973) point out, $r^{(e_r-1)}$ and $X^{(e_x-1)}$ will become a source of systematic error in the stochastic term in the estimating equation and consequently the estimate of the true income-elasticity would be biased. The bias will be in an upward direction if (e_r-1) and (e_x-1) are positive. If these are negative, there will be a downward bias.

One way of abstracting from the problem of the effects of changes in tax-rates which get built into the predictions based on historical trends in tax-revenues is simply to use historical trends only in the tax-base. This is the approach Mushkin and Lupo (1967) favour. In this case, income-elasticity is estimated with reference to the tax-base as in the following equation.

(2.9) $$\log X = \log A' + e'_y \log Y$$

Projections can then be made by first predicting the tax-base (X) and then multiplying it by an effective or an average tax-rate.

To distinguish the elasticity-estimate e'_y in (2.9) from e_y in (2.2), the former is sometimes called the implicit elasticity while the latter is called explicit elasticity.

One further step is to consider partitioning the income-elasticity into its implicit rate and base components, rather than using an effective rate for a given year. The base-elasticity of revenue can be estimated from the following equation.

(2.10) $$\log R = \log A_1 + e'_x \log X$$

Projections can now be made by using (2.10) in conjunction with (2.9) which gives the reduced-form

(2.11) $$\log R = (\log A_1 + e'_x \log A') + e'_y e'_x \log Y$$

The product of the two component elasticities e'_y and e'_x would give the income-elasticity of tax. However, this product will not equal the direct empirical estimate of income-elasticity unless the base and income are perfectly correlated. Where this is not expected to be the case, partitioning may be useful.

A number of difficulties can finally be noted with the approaches outlined above. First, there is the assumption of constant elasticity. Revenue forecasts based on historical estimates of elasticities assume that this elasticity will continue to be the same in the forecasting period as in the sample period. It is in the nature of predictions that something is not allowed to vary. As Theil[1] (1961) says, '...it can be maintained that predictions ... are generated by means of the assumption that something remains constant; the constancy of this "something" is the theory used in the formulation of the prediction.' Whether it is the elasticity coefficient, or the flexibility coefficient or some other such variables which is constant is a question related to the form of the tax-income relationship. It is not necessary that the log-linear form, which implies a constant elasticity, will always describe the tax-income relationship best. This question should remain open until different forms are tried on actual data.

Second, in most of the constant-elasticity models presented above, the only exogenous variable is income. It is the dependent variable which has kept changing from unadjusted revenue to adjusted revenue and tax-base. However, income is not the only factor affecting tax-revenues, and other variables need to be considered.

The exception to this is the model where the tax-rate is treated as an independent variable. This method is useful when there is a single or only a few rate-classes for an individual tax. However, for a tax with multiple rates, it is not clear as to which or how many rates should be used as exogenous variables for deriving meaningful results. In addition to tax-rates and income, of course, one can still search for other relevant exogenous variables.

One way of abstracting from the problem of multiple tax-rates is to use the forecasting model where revenue data are adjusted

---

1. p. 18

for changes in tax-rates, if such an adjustement can be done meaningfully. Even if the adjusted revenue data are available, the predictions generated with the help of these will not be directly useful for tax-policy analysis. These will be predictions based on a constant rate-structure and they will not tell how revenues will vary if rates are manipulated.

Finally, all these models are too restrictive even under partial equilibrium framework as they do not allow for interaction even within the tax-sector of the economy. In other words, these models are geared to forecasting revenues for individual taxes separately and in such a manner that the effect of changes in one tax-yield on another is completely ignored.

2.2 Revenue Adjustment Techniques

The attempt to adjust tax-revenues so that the effect of discretionary changes in tax-rates in the sample period are obliterated is geared towards the prediction of 'automatic' growth in tax-revenues. This is a conditional prediction which assumes that the rate-structure of the year to which revenues are adjusted would prevail in the forecasting period. In other words, it predicts what the revenues would be if such a structure prevailed. As noted earlier, the knowledge of automatic growth in tax-revenues may be desirable; however, the reliability of this knowledge, if it is derived from an adjusted revenue series, depends on the adequacy of the methods used for such adjustments.

Basically, there are two approahces to the construction of an adjusted tax-revenue series. The first method recalculates revenues for all years on the basis of the rate-structure of a reference year. This may be called a 'constant structure method'. It is possible to apply this method only when extensively disaggregated data about tax-bases are avaialble. The second method, which may be called a 'cumulative data adjustment method' adjusts for the revenue effects of each year's legislative action. Examples of the application of the first method are Harris (1966), Singer (1970), and Berney (1970). Examples for the second method are Prest (1962) and, with some modifications, Sahota (1961).

2.2.1 A Constant Structure Method

First, one has to choose a reference year. In general, the current year would seem to be the obvious choice, provided data for this year are available. Having chosen a reference year, the method proceeds in two steps: calculation of 'effective' or 'average' rates in the reference year and then an application of these rates to the tax-bases of the remaining years in the sample.

To illustrate let us take the case of a tax on commodities: say, a sales-tax. Suppose year t is chosen as the reference year. Sales tax revenue in this year will be given by

$$(2.12) \qquad R_t = \sum_{i=1}^{n} r_{it}\, C_{it}$$

where C_{it} is consumption expenditure in the ith class and the tth year, and r_{it} is the tax-rate pertaining to this class.

An adjusted revenue series is obtained by applying, for example, the tax-rates of the tth year to the tax-bases of the remaining years. Thus, for the year (t-j), it is given by

$$(2.13) \qquad R_{t-j} = \sum_{i=1}^{n} r_{it} \cdot C_{i,t-j}$$

The same method can be applied for other taxes on commodities and also for taxes on income. What one does, in fact, is to apply the tax-rates of one year to tax-bases for all other years. The application of this method, however, is dependent upon the availability of highly disaggregated data about tax-bases. In addition, even the theoretical validity of this method, especially in the case of taxes on consumption expenditures, is highly suspect.

A set of different tax-rates can be applied to a set of given tax-bases only under the assumption that a change, for example, in sales tax-rates would not itself lead to a change in tax-bases, i.e. consumption expenditures. Since consumption expenditures are a function of prices, and prices in turn are affected by sales-tax rates, the above assumption will be true only when the price-elasticity of demand is zero. This is too drastic a condition to be valid.

The assumption may not be so drastic in the case of direct taxes. It can perhaps be assumed with greater relevance that the taxable income in a given bracket would be independent of a change in income tax-rates than making the corresponding assumption for sales or import taxes.

2.2.2 A Cumulative Data Adjustment Method

This method requires data on (i) tax-yield and (ii) the revenue-effects of discretionary changes in any year in which they are introduced. Since this method does not require extensive data on tax bases, as was the case with the constant rate method, it may perhaps be more easily adopted in developing countries. The method was proposed by Prest (1962). A basically similar method was proposed by Sahota (1961).

It is frequently the case that when changes in tax-rates, allowances and the like are introduced, the effect of such legislative action on tax-revenue in the current year is predicted in the budget--speech or a similar document at the beginning of the fiscal year. It may also be that, in some cases, the corresponding actual effects of discretionary changes may be announced at the end of the year. If such a series of final data is available, it can be directly used. If it is not obtainable, it will have to be estimated from the predictions made at the beginning of the year.[1] Let $\Delta\hat{D}_t$ be the predicted revenue effect of discretionary action in any year t, and $\Delta\hat{R}_t$ be the predicted total change in that year, the difference between the two being due to automatic growth in tax-revenue. An adjusted series of discretionary effects in individual years can now be obtained from

$$(2.14)\qquad \Delta D_t = \frac{\Delta\hat{D}_t}{\Delta\hat{R}_t}\cdot\Delta R_t$$

where ΔR_t is actual total change in tax-revenue in year t. Other adjustments may also have to be made, e.g., when some discretionary changes are applied to only a part of the year. In the end, suppose that we have constructed a series for D_t over the entire sample period (1,2,. . .,T).

Now let R_{it} indicate the tax-yield in the tth year adjusted to the tax structure of the ith year. Taking a series R_t of actual tax-receipts and adjusting it by ΔD_t, we can obtain the tax revenue in any year adjusted to the structure of its previous year. Thus,

1. These are one-year ahead predictions done by the Ministry of Finance or some such body which is responsible for presenting the annual budget in an economy. The method used for making these estimates is rarely made explicit. In general, it may be assumed that the forecasts arise on the basis of the opinions of experts working in different revenue departments of the ministry concerned. These opinions must ultimately be based on their past observations regarding revenue effects of specific discretionary changes.

$R_{11} = R_1$; $R_{12} = R_2 - \Delta D_2$; $R_{23} = R_3 - \Delta D_3$; $R_{T-1,T} = R_T - \Delta D_T$

The series $R_{11}, R_{12}, R_{23}, \ldots R_{T-1,T}$ is not of much interest in itself. Here each revenue data is adjusted to the structure that prevailed in the previous year by discounting the effect of changes in that year. However, the kind of series that we want is one which is adjusted to the structure of one given year throughout. This we can obtain by a method of cumulative ratios. Suppose year 1 is our reference year. Then the tax-revenue for succeeding years will be adjusted to the structure of year 1 by the following adjustments.

(2.15)
$$\begin{aligned}
R_{11} &= R_1 \\
R_{12} &= R_2 - \Delta D_2 \\
R_{13} &= R_{23} \cdot R_{12}/R_2 \\
R_{14} &= R_{34} \cdot R_{13}/R_3 \\
&\text{----------} \\
R_{1,T} &= R_{T-1,T} \cdot R_{1,T-1}/R_{T-1}
\end{aligned}$$

A simple interpretation can be attached to any one term. For example R_{14} is derived by tax-revenue in year 4 with the structure of year 3 multiplied by the ratio of tax-revenue in year 3 adjusted to the structure of year 1 to actual tax-revenue in year 3.

If, rather than taking year 1, we want to take year T, or the current year, as the reference year the same process will have to be applied in reverse, and the series will be derived by a set of decumulative ratios.

Since the data used in this method is primarily one-year ahead government forecasts regarding the revenue effects of discretionary changes, and since the method as to how these forecasts are derived is not always made explicit, the use of this method is too ad hoc. Further, the method assumes that the changes in revenues following discretionary changes are proportional to changes in total revenues. In other words, it is assumed that the elasticity of the system is not affected by discretionary actions.

2.2.3 Prediction of Automatic Growth in Revenues

The construction of an adjusted revenue series may not always be feasible because of data difficulties. It may also not be desirable to do so because of the limiting assumptions involved in the techniques of such adjustments. However, in specific instances, where such adjustments may be feasible and valid, it would be possible to predict automatic growth in individual tax-revenues on the basis of a given tax structure. The formulation of a prediction model would require a search for exogenous factors that affect tax-bases. Since tax-rates are given, growth in revenues will be directly due to growth in tax-bases.

The basic factor affecting tax-bases is, of course, income. However, as Wilford (1965) suggests, it is important to distinguish between an income-growth which is due to an increase in population with a constant per capita income, and that which is due to a rising per capita income with a constant population. The automatic response of revenues to income-growth would be different if the latter was of the first type rather than the second. For illustration, consider the case of a sales tax. The tax-base for this is taxable consumption expenditure which is dependent on, among other things, total consumption expenditure. This, in turn, is dependent on income. Now if income increases such that per capita income is constant but population has risen, the division of the increased total consumption expenditure between the taxed and the untaxed categories will not be affected. In other words they will rise in the same proportion. This is so because without a rise in per capita income a change in the marginal propensity to consume, either in favour of taxable goods or against it, is not visualised. Hence, the response of revenues to income-growth will be proportionate. On the other hand, if income-growth was due to a rising per capita income with a constant population, the marginal propensities to consume would change one way or the other and hence revenue may rise faster or slower than income.

This analysis is true not only for taxes on commodities, but also for taxes on income. In the case of personal income tax, for example, a rise in per capita income with a constant population would move incomes from non-taxable to taxable income-brackets and revenue will rise faster than income. But if the income-growth was due to a rise in population with a fixed per capita income, the distribution of income between taxable and nontaxable categories should not be affected.

To distinguish between these two types of income growth, it has been suggested that either both total income and per capita income or both population and per capita income should be tried as exogenous variables in the regression equation for an individual tax-revenue.

With some variation, individual tax-revenues with given rate-structures can be written down as a function of population and per capita income. In the case of some specific taxes, this, however, may not suffice or may not be useful. For example, in the case of export taxes, the tax-base, i.e. taxable exports, would be dependent on world demand factors and prices. In the case of corporate income-taxes it may be useful to take the number of corporations rather than population, and average income per corporation rather than per capita income, as the appropriate explanatory variables.

Notwithstanding the limitations of the method used for obtaining adjusted revenues series, predictions of automatic growth in revenues cannot be used directly for policy-analysis because they abstract from predicting the effects of discretionary actions. In view of these difficulties one has still to look for a prediction model that can accommodate policy-parameters.

2.3 A Multiple-rate Tax-interaction Model

The main difficulties with the revenue-estimation models described up to now can be summarised as follows. In these models,
(i) the form of the equations have been prejudged;
(ii) interaction among taxes is not allowed; and,
(iii) applications have been limited to cases where a single or only a few tax-rates have prevailed; in other cases, data adjustment methods have been used, the validity of which is suspect.
In view of these difficulties, it seems necessary to formalise a less restrictive model of tax-revenues. First, consider the problem of multiple tax-rates.

2.3.1 Treatment of Multiple Tax-rates

Because of the unsatisfactory nature of working with adjusted tax-revenue data, it was proposed, as in equation (2.5) that the tax-rate should be taken as an independent variable. This is not difficult where, for an individual tax, there is a single or a general tax-rate. However, in developing economies most of the taxes have a complex rate-structure. All the tax-rates cannot be tried as independent regressors since they will be fairly expensive in terms of degrees of freedom. Sometimes, their number may actually far exceed the size of the sample. Then estimation would not be possible. Besides they may be highly collinear among themselves. It is necessary, therefore, to reduce the whole rate-structure to a few parameters which may provide the relevant regressors. One procedure is suggested below.

Consider, for example, a tax on commodities, say, a sales tax. It is levied on consumption expenditures classified into distinct classes with each class having a distinct sales tax-rate applicable to it. Suppose these categories, which may be termed rate-classes, be J in number. Sales-tax revenues in any given year can be written as

$$(2.16) \qquad R_s = \sum_{j=1}^{J} r_j \, C_j$$

where C_j is <u>taxable</u> consumption expenditure in the jth class after exemptions etc. have been taken into account, and r_j is the tax-rate which pertains to this class.

Now suppose that in any given year all the tax-rates for a given tax can be described by two parameters, say, r^o and r^b such that

$$(2.17) \qquad r_j = r^o + r^b . f(j)$$

where f(j) is a function of j, which varies from 1 to J. r^o can be interpreted as a 'basic rate' and r^b as an 'incremental factor'. The interpretation which may be attached to r^o and r^b is that policy-makers, while deciding about any rate-structure, make decisions about a basic rate and a factor by which rates are increased as we move from a low-rate to a higher-rate class of consumption expenditure. Correspondingly, whenever a new rate-structure is introduced, it will have implicit values for r^o and r^b.

It should be possible to estimate r^o and r^b in any given year by fitting a regression-equation on r_j and j with a sample consisting of J values. In many cases it will be found that a simple form such as

$$(2.18) \qquad r_j = r^o + r^b.j$$

is sufficient.

On the supposition that the whole rate-structure of an individual tax in a given year can be described by an equation such as (2.17), the revenue equation (2.16) can be rewritten as

$$(2.19) \qquad R_s = \sum_{j=1}^{J} [r^o + f(j).r^b] C_j$$

$$= r^o \sum_{j=1}^{J} C_j + r^b \sum_{j=1}^{J} f(j)C_j$$

$$\text{or,} \qquad R_s = r^o. C_s + r^b \sum_{j=1}^{J} f(j)\, C_j$$

where C_s is the total consumption expenditure subject to sales tax.

The formulation of the revenue equation in this manner helps to distinguish between (i) the effect of the total tax-base on the tax-yield through the first term on the right-hand-side of (2.19), and (ii) the effect of the distribution of tax-base into different rate-classes through the second term. In other words, the first term reflects the distribution of total expenditure between taxable and non-taxable coategories; and the second term, that between different categories of taxed consumption expenditures.

In case equation (2.18) is found to be appropriate for describing the rate-structure of a given tax, the revenue equation can be written as

$$(2.20) \qquad R_s = r^o C_s + b \sum_{j=1}^{J} j\, C_j$$

In this case the C_j's are weighted simply by j's which are 1,2,3,...,J.

It is clear that the same analysis is also true for taxes other than a sales tax as long as the related revenue equation can be written in a form such as (2.16). Since not only taxes on commodities and services but also income taxes can be written in this form,[1] the above analysis would seem to be one of general applicability.

2.3.2 Search for Exogenous Variables

In addition to the two tax-rate parameters r^o and r^b, one can now look for exogenous variables that influence (i) total tax-bases, and (ii) their distribution among various rate-classes. This search will have to be geared to individual taxes separately, but there are two variables, viz., population and per capita income, which can be accorded a more general treatment.

Consider again, as an example, the case of sales tax revenue. The base of sales tax, i.e. taxable consumption expenditure, is affected by income. As noted in section (2.2.3), the influence of income can be divided into two parts; that of population and that of per capita income. If growth in income is due to a rising population and a constant per capita income, the pattern of consumption will not be affected and hence the distribution of consumption expenditure between the taxed and the untaxed categories, and that between different rate-classes, will not be affected. Hence the effect on revenue yields will be due to an increase in the aggregate consumption expenditure rather than the distribution of this increase between different rate-classes. On the other hand, if the rise in income results from an increase in per capita income with population unchanged, it is likely that people will spend comparatively more on goods which belong to higher rate classes. Thus, in this case, the influence of an increase in income is traceable to a changed distribution of consumption expenditure in favour of (i) taxed goods as compared to untaxed goods, and (ii) higher-rate goods as compared to lower-rate goods.

1. See section (1.7) in Chapter 1.

It is hypothesised (Legler and Shapiro, 1968) that in developed countries the 'per capita' effect may dampen sales tax revenues as it may lead to an increase in expenditure on non-taxables (e.g. services). Hence, sales tax revenue may increase faster if the rise in it is due to a rise in population rather than one in per capita income. For developing countries probably the reverse is true. As per capita income increases, consumption shifts from low-rate to high-rate classes. It is typically the case that high-rate goods are luxuries and comforts for which people in developing countries have a 'pent-up' demand. When per capita income increases, consumption of these commodities also increases.

To distinguish between these two types of revenue-effects of increased incomes, population and per capita income may be used as independent variables for taxes on commodities. These two variables are also relevant for taxes on income. With a personal income tax, for example, an increase in per capita income with a constant population will increase taxable incomes proportionately more than nontaxable incomes and move these towards higher-rate brackets. On the other hand, an increase in population with per capita income unchanged would proportionately increase total taxable income while leaving the pattern of distribution of income with reference to taxed and untaxed categories, and within the taxed categories, unchanged.

Thus, population and per capita income would seem relevant exogenous variables for most taxes. But the analysis will have to be augmented by specific considerations in the case of each tax. Some of the major taxes commonly used in developing economies are discussed below.

(i) <u>Sales Taxes</u>

For a tax on commodities like a sales tax, in addition to income changes, the influence of movements in prices has also to be considered. Movements in the relative prices of taxed versus untaxed goods will affect the distribution of consumption expenditure between these two categories. This relative price, in turn, will depend on (i) the relevant price indices before the tax rate changes, and (ii) the actual tax rate changes. The relative price of taxed versus untaxed goods may be approximated by the ratio of a price index of major taxed items to that of major untaxed items. A one year lag may be introduced in the index of this ratio to approximate the relative price before the tax rate change in any given year. This lagged index may be denoted by P_{-1}. The influence of sales-tax rates in a multiple-rate system will be reflected by r_s^o and r_s^b where these are the 'basic' sales tax-rate and its incremental factor, in accordance with the considerations previously introduced. In addition, since consumption expenditures are made after personal income taxes have been paid or allowed for, total personal income-tax revenue will also be a factor affecting sales tax-revenue.

A revenue estimation equation for sales tax revenue (R_s) can now be written down as

$$(2.21) \qquad R_s = R_s(N, y, r_s^o, r_s^b, P_{-1}, R_y)$$

where N is population, y is per capita income, and R_y is personal-income tax revenue.

In accordance with the reasoning presented in the previous section, the relationship between R_s and N and y is expected to be positive. The expected sign of the relation between R_s and r_s^o and r_s^b is a composite of two influences in opposite directions. First, since r_s^o and r_s^b enter as <u>multiplicative</u> factors as in equation (2.19), there will be a positive relationship. Second, as they also influence the taxable consumption expenditure by altering the relative price of taxed vs. untaxed goods, there will be a negative relationship, i.e. the higher the sales tax-rate parameters, the greater will be the shift away from taxable items and hence a reduction in the tax-base. The net result will depend on the strength of one type of influence as compared to the other.

The influence of P_{-1} is expected to be positive. The higher this index is, the smaller will be the shift away from the taxed goods following a change in the sales tax rates. The influence of R_y will, however, be negative. The greater the sums paid out in personal income taxes, the smaller will be the sums available for consumption expenditure.

(ii) Excise Taxes

Since excise tax is a tax on domestically produced commodities, it is subject to an analysis analogous to that of a sales tax except that a different price index should be chosen to stand for the index of the price of taxed vs. untaxed goods. This price index should be made with reference to major items subject to excise tax. If we write this index with a year's lag as P'_{-1}, we have a revenue estimation equation for excise tax revenue as follows.

$$(2.22) \qquad R_e = R_e(N, y, r_e^o, r_e^b, P'_{-1}, R_y)$$

where r_e^o is the 'basic' excise-tax rate and r_e^b its 'incremental factor.'

(iii) Import Tax

Here again population and per capita income effects will be important. An increase in income which is due to a rise in population with per capita income unchanged will mean a symmetric increase in expenditure on both imported and domestic goods. But if the increase in income was due to an increase in per capita income with population held constant, it may change the pattern of expenditure in favour of imported goods as a high proportion of these, in a developing economy, are generally luxuries and high-cost capital goods.

Further, we should bring in variables that govern the distribution of expenditure (i) between imported goods and domestic goods, and (ii) between taxed and untaxed imports. The first kind of distribution should be governed by the relative price of imported goods versus domestic goods. This can be obtained by the ratio of unit value imports to domestic prices. Suppose we use an index P* to represent the ratio of the unit value of imports to domestic wholesale prices. The higher this index, the lower will be the demand for and hence the expenditure on imports. Thus, we expect a negative relationship between import tax-revenue R_m and p*.

The distribution of import-expenditure between taxed and untaxed imports in any given year will depend on (i) the price of imported goods which are taxed relative to the price of untaxed imported goods before the tax-rate changes and (ii) the tax-rate changes. If indices of the prices of major taxed and untaxed imports can be prepared, an index of the ratio of these two, lagged by a year, would give an approximation of factor (i) above. For the second, the 'basic' rate of import tax r_m^o and its incremental factor r_m^b may be used. The relation of R_m with p_{-1}^m should be positive, i.e. the higher p_{-1}^m, the smaller should be the movement <u>away</u> from taxed imports after the tax-rate changes are introduced.

The influence of r_m^o and r_m^b will again be a composite of two influences: one with a positive effect on R_m and the other, a negative effect. The positive effect stems from the use of r_m^o and r_m^b as multiplicative factors in a revenue-generation equation like (2.19). The negative effect arises from the fact that the higher the tax-rate parameters, the larger will be the import prices after the tax-rate changes relative to the prices before the changes, and hence the smaller will be the expenditure on imports. However, for import taxes in developing countries, many tax-rates become so high as to be prohibitive. In these cases, the negative influence of tax-rates may be considered to outweigh the positive effects, especially as far as consumption goods are conçerned. Still, the balance of signs may go either way and would depend upon the given empirical situation.

In addition one might like to consider taxes on income as having a dampening effect on expenditure on imports and hence include both personal income-tax revenue (R_y) and corporate income-tax revenue (R_c) in the revenue estimation equation for import taxes. The latter term is included so as to primarily take account of its influence on capital imports by corporations. Both R_y and R_c will be expected to have a negative influence on R_m. Thus, finally we can write a revenue-estimation equation for import-taxes as follows.

$$(2.23) \qquad R_m = R_m\,(N, y, p_{-1}^m, p^*, r_m^o, r_m^b, R_y, R_c)$$

Since there are several dependent variables, the loss of degrees of freedom will be substantial if the sample size is small. In such a case, some of these variables may have to be dropped.

(iv) Personal Income Taxes

Again, growth in population should be taken as a factor symmetrically augmenting both taxed and untaxed incomes. In contrast, an increase in per capita income should be taken as a factor which moves incomes from untaxed to taxed brackets and from low-rate to high-rate brackets. Both the symmetric and the asymmetric influences increase the personal income-tax revenue and hence the expected sign of the relationships should be positive. The tax-rate parameters can again be represented by r_y^o and r_y^b. Since, for this tax, these are only multiplicative factors, their influences will be positive. The revenue estimation equation is then given by

$$(2.24) \qquad R_y = R_y\ (N,\ y,\ r_y^o,\ r_y^b)$$

where r_y^o is the 'basic' income-tax rate and r_y^b, its 'incremental factor.'

(v) Corporate Income Taxes

The type of exogenous variables which applied to personal income taxes are relevant here also. However, rather than population, the number of corporations, and, rather than per capita income, the income per company are the more relevant variables. Denoting the number of companies by N^*, and the income per company by y^*, corporate-income tax revenue can be estimated by

$$(2.25) \qquad R_c = R_c\ (N^*,\ y^*,\ r_c^o,\ r_c^b)$$

where r_c^o and r_c^b are respectively the 'basic' corporate tax-rate and its 'incremental factor'. The influence of all the exogenous variables on R_c will be positive by reasoning analogous to that set out in the case of personal income tax.

(vi) A consolidated Model

Suppose there are five taxes in an economy, viz., a sales tax, an excise tax, an import tax, a personal income tax and a corporate income tax. These taxes, in fact, cover the major indirect and direct taxes in a developing economy. Where there are other taxes playing a significant role in a particular economy, individual equations for these will be needed. The remaining minor taxes can be grouped into just one category of 'other taxes'. Since the importance of this category would be small in terms of its contribution to total revenue, we may predict this component of revenue by a simple extrapolation. A highly simplistic model would be to take the current value of this remainder term just as a function of its last year's value. Now, the total revenue (R) in any given year can be written as

$$(2.26) \qquad R = R_s + R_e + R_m + R_y + R_c + R_r$$

where R_r is the remainder term under 'other taxes'.

For each of the terms in identity (2.26) an equation can be written down.

$$(2.27) \qquad \begin{aligned} R_s &= R_s(N,\ y,\ r_s^o,\ r_s^b,\ P_{-1},\ R_y) \\ R_e &= R_e(N,\ y,\ r_e^o,\ r_e^b,\ P'_{-1},\ R_y) \\ R_m &= R_m(N,\ y,\ P^m_{-1},\ p^*,\ r_m^o,\ r_m^b,\ R_y,\ R_c) \\ R_y &= R_y(N,\ y,\ r_y^o,\ r_y^b) \\ R_c &= R_c(N^*,\ y^*,\ r_c^o,\ r_c^b) \\ R_r &= R_r(R_{r,-1}) \end{aligned}$$

In this model the following exogenous variables are used in addition to the tax-rate parameters.

N = population

y = per capita income

N^* = no. of companies

y^* = average corporate income per company

p = price index for major sales tax items

p' = price index for major excise tax items

p^m = unit value of imports

p^* = unit value of imports upon domestic prices

The subscript '-1' refers to one year's time-lag.

Actual prediction will require (i) a judgement regarding the form of the equation, (ii) the estimation of the coefficients, and (iii) the use of independently predicted values of exogenous variables. On the a priori reasoning set out in the previous section, a table of expected signs can be prepared at this stage.

Table of Expected Signs

Variables	N	y	N*	y*	p_{-1}	p'_{-1}	p^m_{-1}	p*	R_y	R_c
R_s	+	+			+				-	
R_e	+	+				+			-	
R_m	+	+					+	-	-	-
R_y	+	+								
R_c			+	+						

The relationship of each individual tax to its rate parameters (r^o, r^b) is expected, in general, to be positive when it as assumed that the positive multiplicative effect is likely to outweigh the negative effect on tax-bases. But it cannot be said a priori that this would always be the case.

2.4 Some Additional Considerations

Developing economies characteristically defy formalization of the behaviour of their economic variables into a typical model framework. An attempt to build a revenue estimation model which will uniquely fit the set of economies classified under 'developing' will not be very rewarding. Hence, the model developed in the previous section can only serve as a frame of reference. It will need to be augmented by additional considerations for any given empirical situation.

At a more theoretical level, however, it is possible to anticipate some typical problems and suggest techniques for dealing with them.

2.4.1 Ad Hoc Events and Use of Dummy Variables

The initiation of a development process is frequently characterised by important once-and-for-all changes in the economy. It is not always possible to convert these changes into meaningful economic variables so as to catch their influence in a classical regression framework. But prediction models have to accommodate changes of this nature. A devaluation of the domestic currency within the sample period, a major change in the system of adiministration and collection of revenues, a conversion of specific tax-rates into ad valorem ones are events which will have considerable impact on the yield of taxes. Such events can be represented by introducing dummy variables into the estimating revenue equations.

Dummy variables are designed to take account of situations where a 'step function' relation between the dependent and one or more of the independent variables is visualized.

Step functions may typically look like the ones shown in Diagram 2.1.

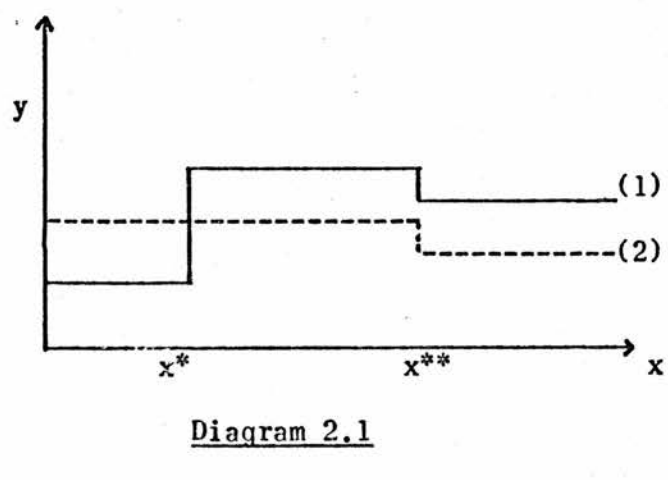

Diagram 2.1

For the step function (1), the break-off points are x^* and x^{**}. For function (2) the break-off point is x^{**}. These break-off points may be things like the year in which a devaluation took place, the year in which a Finance Commission gave its award, the year in which a different political party took over administration, and so on.

Provided that the effect of these events is such that only the 'level' of the dependent variable is changed at the break-off points, it is possible to represent these cases by a set of dummy variables.

For example, in case (1) above, we can define the regressor variables as

$$x_1 = \begin{cases} 1 \text{ if } x \leq x^* \\ 0 \text{ otherwise} \end{cases}$$

$$x_2 = \begin{cases} 1 \text{ if } x^* < x \leq x^{**} \\ 0 \text{ otherwise} \end{cases}$$

$$x_3 = \begin{cases} 1 \text{ if } x > x^{**} \\ 0 \text{ otherwise} \end{cases}$$

Similarly, case (2) can be represented by

$$x_1 = \begin{cases} 1 \text{ if } x \leq x^{**} \\ 0 \text{ otherwise} \end{cases}$$

$$x_2 = \begin{cases} 1 \text{ if } x > x^{**} \\ 0 \text{ otherwise} \end{cases}$$

Suppose we are trying to predict tax-revenue from import duties, where a devaluation of the domestic currency has taken place, say, in the third year of the sample. If an additional explanatory variable y (income) is being used, the data matrix for independent variables, using dummy variables to define pre- and post-devaluation periods, will look like the following.

$$\begin{bmatrix} 1 & 0 & y_1 \\ 1 & 0 & y_2 \\ 1 & 0 & y_3 \\ 0 & 1 & y_4 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0 & 1 & y_n \end{bmatrix}$$

The estimating equation can be written as

(2.28) $$R_m = \beta_1 x_1 + \beta_2 x_2 + \beta_3 y + \epsilon$$

where x_1 and x_2 take values equal to zero and one as shown in the matrix. The first term in the right-hand-side of (2.28) will measure the predevaluation intercept, and the second, the post-devaluation intercept.

Least squares estimates of β_1 and β_2 can be obtained with appropriate assumptions about ϵ.[1] It should be noted that in this scheme there is no place for a constant term for the entire sample period. This is so because if there were to be another variable $x_3 = 1$ for the entire sample period, a linear dependence between the regressors (i.e., $x_1 + x_2 - x_3 + 0.y = 0$) will produce singularity in the estimation matrix.

This creates a problem in using a computer algorithm which already produces an intercept. If one continues to use this, the way out is to rewrite (2.28) in the following form.

(2.29) $$R_m = \gamma_1 + \gamma_2 x + \beta_3 y + \epsilon$$

where x=0 in each predevaluation year and 1 in each post-devaluation

---

1. For a detailed analysis of dummy variables and their least squares estimates see e.g. Johnston (1963, 1972) and Goldberger (1964).

year.

The pre-devaluation intercept (x=0) will be γ_1 (= β_1). The post-devaluation intercept (x=1) will be $\gamma_1 + \gamma_2$ (= β_2). This is to say that γ_2 is equivalent to $\beta_2 - \beta_1$, i.e. the difference between the post- and pre-devaluation intercepts. When this scheme is used for estimation, care should be exercised in conducting significance tests on the coefficients. In particular, the significance of the post-devaluation intercept should be judged with respect to $\gamma_1 + \gamma_2$ and not with respect to γ_2 alone.[1]

The use of dummy variables should prove convenient in revenue estimation where ad hoc events such as the ones noted above are encountered. More than one set of dummy variables can be used in regression analysis and if it is desired, interaction among them can also be permitted. The basic limitation in their use is that the functions behind them remain flat and that they are expensive in terms of degrees of freedom.

2.4.2 Multiple Tax Rates and Principal Components Analysis

In practice it will be found that in a developing economy, especially in the case of indirect taxes, there are so many classifications of goods, each with a different rate of tax, that their representation within the framework stipulated earlier is not a feasible scheme. It is not only that different commodities may be subject to different tax-rates but that the same class of commodity may be subject to various classifications for the application of different tax-rates. The simultaneous use of these tax-rates as independent variables is precluded because of the limitation imposed on the degrees of freedom

---

1. For an extensive treatment of this point, see Johnston (1963, 1972).

and a possible problem of multicollinearity.

One way out in these situations is to do a principal component analysis on the tax-rates. This analysis transforms a set of k variables to a new set of k variables which are linear combinations of the original set and which are pairwise uncorrelated. Further, the new variables are so arranged that the first variable has the maximum possible variance, the secon, the maximum possible variance among those uncorrelated with the first and so on.[1] In this scheme, the first few of the transformed variables will explain a substantial portion of the variation of the original variables. These may be retained for a regression analysis while the remaining principal components can be dropped. Thus, the number of regressors will be reduced, and there will not be a problem of multicollinearity.

Some tests[2] are available which give an approximate idea of the number of components to be retained for the regression analysis. Occasionally a choice based on observation may prove to be adequate.

It is desirable to derive the principal components for the tax-rates alone leaving the other regressors, such as income and prices. Then it will be easier to attach some kind of interpretation to the principal components. If these are computed for variables

---

1. For a matrix X of n observations on k variables, the principal components are given by an nxk matrix Z, such that

$$Z = XA$$

where A is a (lxk) matrix. The elements in the X matrix are standardized. Each element in A is a latent vector of X'X corresponding to one of its latent roots. The latent vectors are derived so as to satisfy the conditions of orthogonality and successive maximum variation. The first vector a_1 corresponds to the largest root, the second vector a_2 correspons to the second largest root and so on.

2. For the derivation of the principal components, and for tests regarding the choice cf the number of components to be retained in a regression analysis, see Johnston (1963, 1970).

measured in different units, it will not be possible to interpret them. Since it is possible to use the principal components along with other explanatory variables in a regression analysis, it seems apt that principal components of tax-rates should be used in conjunction with other regressors.

2.4.3 Tax Demand and Tax Revenues

The revenue-generation model such as $R = \sum_i r_i x_i$ is, in fact, a tax-demand model and not a tax-revenue (or tax-collection) model. The tax-rates applied to tax-bases represent the demand for revenue collection in any given year. The actual tax collection in a year is generally not the same. The actual collection is the sum of (i) past arrears, (ii) a part of future anticipated tax-demands, such as taxes paid in advance, and (iii) a substantial part of the current years' tax demand.

If one writes R_t for revenue collected in year t and D_t for revenue demanded in this year, it can be said that

$$(2.30) \qquad R_t = aD_t + bD_{non-t}$$

The part of this year's tax-demand which spills over in other years whether in future or in the past is $(1-a)D_t$. The part of other year's tax-demand which augments this year's tax-revenues is bD_{non-t}. If these two elements balance out, R_t will be a perfect proxy for D_t, and an analysis can be conducted with R as the dependent variable. If the difference between the two elements is expected to be more than marginal, the proper thing is to hypothesise two relations, viz.,

$$(2.31) \qquad D_t = \sum_i r_{it} x_{it}$$

and

$$(2.32) \qquad R_t = f(D_t, D_{t+i}, D_{t-i})$$

Predictions will then have to be made from the reduced-form equations for R_t, which will be otained by substituting estimated tax-demand equations in the estimated relationship between tax-demand and tax-revenue. Occasionally, a simple relationship between R_t and D_t will suffice. In other cases, a decision about a lag-structure in D_t will be required.

CHAPTER 3

EVALUATION OF FORECASTS

Forecasts are evaluated against actual data and in relation to other forecasts. Such comparisons help in assessing the reliability of forecasting models and in choosing between alternative techniques and models.

Forecast-evaluation techniques have been developed in a wide context. In general they are as applicable to revenue forecasts as to any other set of forecasts. Exercises in forecast-evaluation generally assume that actual values or realizations of the predicted variables are available. As a first step, the size and nature of forecast errors, i.e. the deviations of a prediction from a corresponding realization, are described. Techniques which serve these ends can be grouped under 'absolute accuracy analysis.'

However, these techniques do not offer a sufficient means of comparing between different forecasting techniques when forecasting periods or predicted variables differ. Hence, the predictions are sometimes compared with some 'benchmark' forecasts so as to form an idea about the 'relative accuracy' of predictions. The 'benchmark' predictions may originate from another technique or source.

Although considerations here are restricted to absolute and relative accuracy analyses, it should be noted that by themselves, these are not a sufficient guide to the value of a forecast. A forecast must ultimately be evaluated in relation to its purpose, which, in the end, is to aid decision-making processes wherever uncertainty is involved. One needs, therefore, to compare the quality of decisions that would be made when predictions are available and when they are not. Such an analysis can be done with the help of 'loss' functions (Theil, 1961). There are considerable practical difficulties in defining and estimating loss functions. As a first approximation, it may be assumed that the reduction in loss due to any set of predictions can be adequately indicated either by their deviations from realizations or by a comparison with some 'benchmark' predictions.

3.1 Absolute Accuracy Analysis

This consists of devising methods to describe prediction errors assuming that two time series with finite and homogenous time intervals, one relating to predictions and the other relating to realizations, are available. Let these series respectively be P_t and A_t, t varying from 1 to n. There are various graphical and statistical ways which can be used to describe the error $U_t = P_t - A_t$. First, however, a distinction between levels of variables and changes in levels needs to be made.

3.1.1 Levels and Changes

Most of the descriptive measures used in the literature relate to changes rather than levels. Let the predicted and realized changes for period t be indicated by ΔP_t and ΔA_t respectively.

When models predict the levels of economic variables, there are two possible ways of obtaining predicted change data: (i) successive differences between predicted levels i.e. $\Delta P_t = P_t - P_{t-1}$;

(ii) differences between the predicted level for a period and the actual level for the previous period, i.e. $\Delta P_t = P_t - A_{t-1}$.

The latter approach is used in ex-post evaluations. Here, $\Delta P_t - \Delta A_t = (P_t - A_{t-1}) - (A_t - A_{t-1}) = P_t - A_t = U_t$, which is the error in levels. In the other case,

$\Delta P_t - \Delta A_t = (P_t - P_{t-1}) - (A_t - A_{t-1}) = (P_t - A_t) - (P_{t-1} - A_{t-1})$
$= U_t - U_{t-1} = \ U_t$, which is the error relating to changes.

ΔA_t is normally taken to be equal to $A_t - A_{t-1}$. In some cases, when the realized value of A_t is not available, an estimate of it, say, A^*_t, is used.

3.1.2 The Prediction-Realization Diagram

A plot of predicted and realized changes on the cartesian axes offers a good visual description of forecasting performance. The axis assigned to the predicted or realized change depends on what is to be observed. Predicted changes are plotted along the Y-axis if we want to observe how predicted changes are distributed given the realized changes. Realized changes are plotted along the Y-axis if we want to know the distribution of realized changes given the predicted changes. The two cases are shown in diagrams 3.1 and 3.2. The upward sloping 45^0 line is the line of perfect forecasts (LPF).

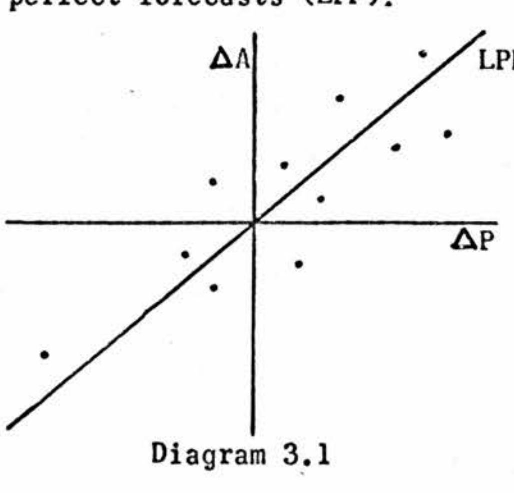

Diagram 3.1

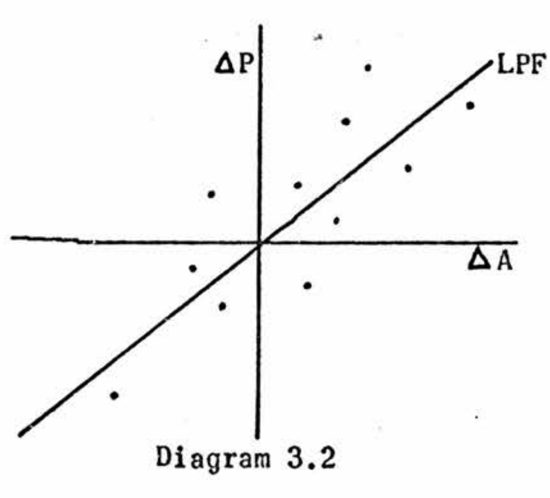

Diagram 3.2

A prediction-realization diagram (Theil, 1961, 66) is obtained by rotating diagram 3.2 where realized changes are represented on the X-axis, such that the line of perfect forecasts becomes a horizontal line. This is shown in diagram 3.3.

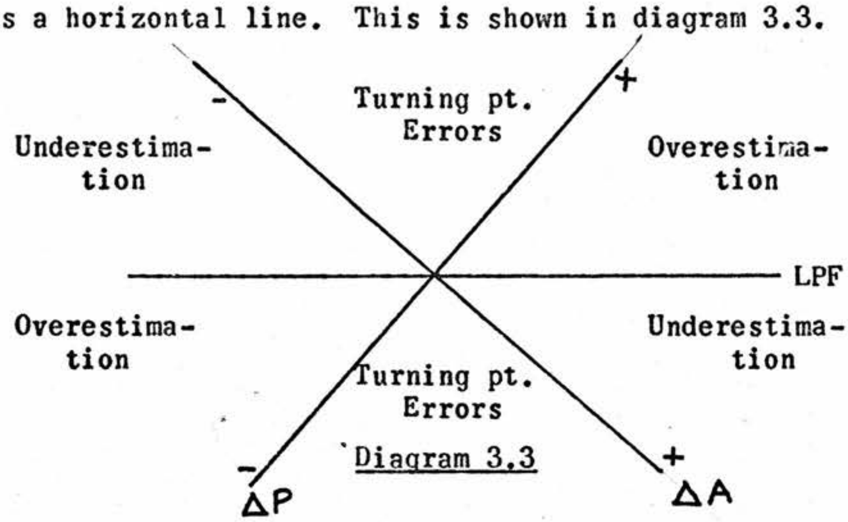

Diagram 3.3

Observations about the correct prediction of the direction of change, under- and over-estimation, and turning point errors can be made in all three diagrams.

The direction of change will have been predicted correctly if the predicted change has the same sign as the realized change. This is so for points in the first and third quadrants of Diagrams 3.1 and 3.2 and areas on the right and on the left of the P-R diagram.

If the direction has been predicted correctly, but the forecasts are not perfect, the change must have been either over-estimated or under-estimated. Underestimation of changes occurs below the L.P.F. in the first quadrant of Diagram 3.1 and above it in the third quadrant. In Diagram 3.2, it is above the L.P.F. in the first quadrant and below it in the third quadrant. The respective remaining areas in the first and third quadrants of these diagrams apart from the L.P.F. itself contain cases of overestimation. In the P-R diagram, areas relating to under- and over-estimation are indicated.

The direction of change will not have been predicted correctly for points in the second and fourth quadrants of Diagrams

3.1 and 3.2 or in the areas above or below the origin in the P-R diagram. These are cases of turning point errors which may be of two types. A turning point may be predicted where it did not occur, or it may not have been predicted where it did occur.

Diagrammatic representations such as 3.1, 3.2 and 3.3 help to distinguish between different types of error and thus aid the analysis of forecasting errors. But for adequately comparing different forecasting performances, statistical measures are needed.

3.1.3 Statistical Measures

Given a series P_t of predictions and A_t of realizations for a homogenous period, a number of statistical measures can be used to index the prediction performance. Some such measures are given below.

Average absolute error is defined as

$$\sum_{t=1}^{n} /P_t - A_t/ = \sum_{t=1}^{n} /U_t/$$

The mean square prediction error (M_p), which is the most popular of such measures is given by

$$M_p = \frac{1}{n} \sum_{t=1}^{n} (P_t - A_t)^2 = \frac{1}{n} \sum_{t=1}^{n} U_t^2$$

The reasons for its attractiveness as a measure of forecasting accuracy are the same as those for variance in statistical analysis, viz., its statistical and mathematical properties. Sometimes its underroot, called the root-mean-square-error, is used. It is given by

$$\sqrt{\frac{1}{n} \sum_{t=1}^{n} (P_t - A_t)^2}$$

All these measures have a value zero for perfect forecasts, i.e. $P_t = A_t$ for all t. They do not have a finite upper limit and they are all positive. Thus, the higher the value they assume, the worse the forecast must be.

The mean square prediction error has been used in two measures defined as inequality coefficients. In practical applications, these have been much favoured. The first inequality coefficient (Theil, 1961)[1] was defined as

$$U_1 = \frac{\sqrt{\sum(P_t - A_t)^2}}{\sqrt{\sum P_t^2} + \sqrt{\sum A_t^2}}$$

This coefficient varies between zero and unity. It is indeterminate for the trivial case when all the predictions and realizations are zero. It is zero for perfect forecasts. The higher its value, the lower is the indicated quality of the forecast.[2]

A second inequality coefficient has been defined (Theil, 1966) as

$$U_2 = \sqrt{\frac{\sum(P_t - A_t)^2}{\sum A_t^2}}$$

The alteration occurs only in the denominator which indicates the search for an appropriate unit of measurement. For perfect forecasts again, the value of the coefficient is zero.

If this coefficient is written in terms of changes rather than levels, i.e.

---

1. Although Theil has defined the coefficients specifically for changes rather than levels, they can also be written, as above, for levels.

For some measure of dissatisfaction with the adequacy of inequality coefficients as devices for forecast evaluation, see Granger and Newbold (1973).

2. For an interpretation of the meaning of an inequality coefficient equal to 1 see Theil (1961).

$$U_2' = \sqrt{\frac{\sum(\Delta P_t - \Delta A_t)^2}{\sum \Delta A_t^2}}$$

it will take a value equal to unity for a no-change extrapolation, i.e. $\Delta P_t = 0$ for all t. Whether defined for changes or levels, this coefficient has no finite upper bound.

Another measure for accuracy of forecasts is the simple correlation coefficient between P_t and A_t, i.e.

$$r = \frac{\sum(P_t - \bar{P})(A_t - \bar{A})}{\sqrt{\sum(P_t - \bar{P})^2 \sum(A_t - \bar{A})^2}}$$

The difficulty with this measure is that it is independent of the origin and the unit of measurement. The unit does not create much of a problem as P_t and A_t would normally be written in the same units. But the question of origin does create a problem. This measure would not distinguish between situations where the levels of P_t are consistently higher or lower than the levels of A_t. This kind of discrepancy is a systematic source of error and is called 'bias'. This concept is taken up in the next section. The correlation coefficient may be an adequate measure for 'unbiased' forecasts and forecasting accuracy is judged by the deviation of r from unity.

3.1.4 Systematic Errors in Forecasts

A regression of prediction on realizations or vice versa can provide valuable insight into the nature of prediction errors. The two relationships can be written as

(3.1) $$P_t = a + bA_t + e_t$$

(3.2) $$A_t = \alpha + \beta P_t + v_t$$

where e_t and v_t are the respective residuals.

A value for a or α equal to zero means that the relevant regression line passes through the origin, and a value for b or β equal to one implies its coincidence with the line of perfect forecasts. The non-zero values of a and α and the non-unity values of b and β will be seen to be sources of systematic errors in the forecast. For this, we need to concentrate on just one regression line.

Preference has been shown for the regression of A_t on P_t (Mincer and Zarnowitz, 1969) for two reasons: first, P_t's are, by definition, the predictors of A_t's; and secondly, they are available before the realizations. A regression line of this type is drawn in diagram 3.4.

The mean point ($\bar{P}$, $\bar{A}$) is seen not to lie on the L.P.F. This is a source of systematic 'bias' and can be removed by shifting the regression line until the mean point lies on the L.P.F. As it is desirable for the mean point to be on the L.P.F. so also it is 'intuitively' desirable that the whole regression coincides with the L.P.F. If this is so, the forecast is called 'efficient'. When it is not so, such efficiency can be obtained by rotating the shifted regression line so that it coincides with the L.P.F.

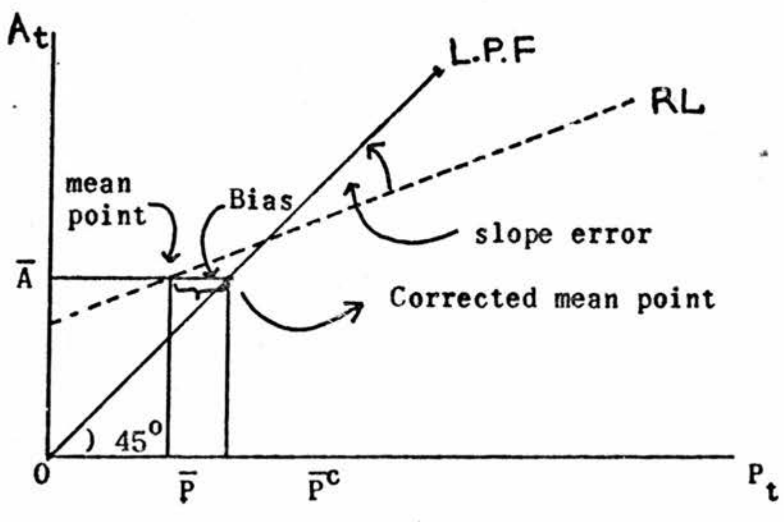

Diagram 3.4

In practice, these two changes are obtained by setting the least squares value of

$$\alpha = \bar{A} - \beta \bar{P}$$

$$\beta = \frac{\sum (P_t - \bar{P})(A_t - \bar{A})}{\sum (P_t - \bar{P})^2}$$

respectively equal to zero and unity.

Once each forecast is multiplied by β and added to a constant α, the new set of forecasts has undergone what has been called an 'optimal linear correction' (Theil, 1961). The two changes can be interpreted by analyzing the mean square prediction error. A decomposition of M_p is given by

(3.3) $$M_p = (\bar{P} - \bar{A})^2 + \sigma_u^2$$

When the forecast is corrected for bias, $\bar{P} = \bar{A}$. This yields a mean square error for the corrected forecast

(3.4) $$M_p' = \sigma_u^2$$

When the forecasts are also made efficient. i.e. $\beta = 1$ in (3.2), it is implied that the forecast error u_t is uncorrelated with the forecast values P_t. In this case the variance of the residuals σ_v^2 equals the variance of the forecast error σ_u^2. Otherwise, $\sigma_u^2 > \sigma_v^2$. Hence, the definition of the efficiency of forecasts has been given as $\sigma_u^2 = \sigma_v^2$.

An unbiased and efficient forecast will have $\alpha = 0$, and

(3.5) $$M_p = \sigma_v^2 = \sigma_u^2$$

In general,

(3.6) $$M_p \geq \sigma_u^2 \geq \sigma_v^2$$

3.1.5 Decompositions of Mean Square Prediction Error

It is possible to decompose σ_u^2 in equation (3.3) in a number of ways. The following decompositions have been proposed in the literature.

$$(3.7) \qquad \sigma_u^2 = (1 - \beta)^2 \sigma_P^2 + (1 - r^2)\, \sigma_A^2$$

$$(3.8) \qquad \sigma_u^2 = (\sigma_p - \sigma_A)^2 + 2(1 - r)\, \sigma_p \sigma_A$$

$$(3.9) \qquad \sigma_u^2 = (\sigma_p - r\,\sigma_A)^2 + (1 - r^2)\, \sigma_A^2$$

Here, β is the regression coefficient in the regression of A_t on P_t, r is the coefficient of correlation between P_t and A_t, and σ_p and σ_a are respectively the standard deviations in the predicted and realized series.

These decompositions of σ_u^2 can be introduced in equation (3.3), so as to provide three decompositions of the mean square error. Each form is subject to a relevant interpretation.

In the first case,

$$(3.10) \qquad M_p = (\overline{P} - \overline{A})^2 + (1 - \beta)^2 \sigma_P^2 + (1 - r^2)\, \sigma_A^2$$

The three components on the right hand side have respectively been called the mean, slope, and residual components (Mincer and Zarnowitz, 1969). The first term is zero for $\overline{P} = \overline{A}$. Hence, any positive value for it is due to errors in central tendency. As has already been seen, the deviation between the two averages is indicative of 'bias' in the prediction. Once this systematic error is removed, M_p will be reduced by the value of the mean component.

The second term in (3.10) will assume a value zero for $\beta = 1$. β is the slope of the line of regression of A_t on P_t, and for this reason this term is called the slope component. The slope error is also a systematic error. Its removal will reduce M_p by the value of the second term in (3.10).

A forecast which is made unbiased and efficient in this sense, will have a mean square error equal to the third component. It is, therefore, called the residual component.

Consider now the decomposition of σ_u^2 given by (3.8). Introducing it in (3.3), we have

$$(3.11) \qquad M_p = (\bar{P} - \bar{A})^2 + (\sigma_p - \sigma_A)^2 + 2(1 - r)\sigma_A \sigma_p$$

This decomposition is symmetric in P and A. The first component is the same in all decompositions and subject to the same interpretation. The second and third terms shed more light on the nature of sources of forecast error.

The second term would be zero when the two standard deviations are equal. Any positive value for this term may, therefore, be interpreted as error due to incomplete variation. Errors of this nature arise because of the forecaster's neglect of the causes of fluctuations in the two series. Although the forecaster is not expected to be perfect in this regard, he should be able to reduce this type of error over time.

The situation in the third term has been described by Theil (1961) as more 'hopeless'. This term is zero when r=1 or also when $\sigma_A \sigma_p r = \sigma_A \sigma_p$, i.e. the covariance of predictions and realizations ($r \sigma_A \sigma_p$) takes its maximum value, viz., the product of the two standard deviations. Any positive value for this term is, therefore, due to incomplete covariation. These errors stand less chance of correction.

Following the nature of errors, the second and third terms in equation (3.11) may be called variance and covariance components of the mean square error.

A third decomposition can be derived by rewriting the second term in (3.10). Thus, we have

$$M_p = (\bar{P} - \bar{A})^2 + (\sigma_p - r\,\sigma_A)^2 + (1 - r^2)\,\sigma_A^2 \tag{3.12}$$

The interpretation of the three terms in this case remains the same as in the first decomposition since the second term will be zero only when $r\,\sigma_A/\sigma_p$, which is equal to β, is unity. However, Theil (1966) uses this form to derive certain inequality proportions.

3.1.6 Inequality Proportions

The division of the two decompositions proposed in equations (3.11) and (3.12) by their sum, i.e. $\frac{1}{n}\sum(P_t - A_t)^2$, yields two sets of inequality proportions. These quantities are convenient for indicating the relative contribution of the individual sources of errors as interpreted in the respective decompositions.

Thus, from equation (3.11) the following set of inequality proportions is derived.

$$(3.13)\quad \begin{cases} U^M = \dfrac{(\bar{P} - \bar{A})^2}{\frac{1}{n}\sum(P_t - A_t)^2} \\[2ex] U^s = \dfrac{(\sigma_p - \sigma_A)^2}{\frac{1}{n}\sum(P_t - A_t)^2} \\[2ex] U^c = \dfrac{2(1 - r)\,\sigma_p\,\sigma_A}{\frac{1}{n}\sum(P_t - A_t)^2} \end{cases}$$

Corresponding to the interpretations attached to the numerators, the three proportions have respectively been called bias, variance, and covariance proportions.

Similarly, the second set is given by the following.

$$(3.14)\quad \begin{cases} U^M = \dfrac{(\bar{P} - \bar{A})^2}{\frac{1}{n}\sum(P_t - A_t)^2} \\[2ex] U^R = \dfrac{(\sigma_p - r\,\sigma_A)^2}{\frac{1}{n}\sum(P_t - A_t)^2} \\[2ex] U^D = \dfrac{(1 - r^2)\,\sigma_A^2}{\frac{1}{n}\sum(P_t - A_t)^2} \end{cases}$$

The first proportion in the two sets is the same. The second and third proportions in the second set are respectively called regression and disturbance proportions, thus indicating the nature of the source of error evidenced in their numerators.

It is obvious that the proportions would add up to unity in each case, i.e.

$$(3.15)\qquad U^M + U^S + U^c = 1$$

$$(3.16)\qquad U^M + U^R + U^D = 1$$

The relevance of the study of forecast error in terms of inequality proportions lies in the fact that having identified a source of high relative contribution to the prediction error, efforts can now be concentrated in that particular direction to improve the forecasts.

3.2 Relative Accuracy Analysis

Absolute accuracy analysis can be done when data about realizations is available. Even when this is so, different forecasting methods cannot be ranked with the help of size and characteristics of forecast errors if their prediction horizon and predicted economic variables differ. To produce an unambiguous ranking of forecasting performances, some 'rate of return' criterion is needed. With the help of such a criterion, it may be possible to compare the gains obtainable from the reduction of forecasting error by using one method of prediction rather than another. This would make a comparison of different methods possible even when target dates or predicted variables differ.

This necessitates the use of some kind of gain or loss function which measures the consequences of the reduction in forecasting error. It is difficult to obtain such functions in practice. In their absence, Mincer and Zarnowitz (1969) offer a limited criterion. The gain is measured here simply in terms of the reduction in forecasting error obtained by the forecasting method in question relative to some 'benchmark' method. If the mean square errors in the two methods are respectively called M_p and M_x, the proposed index of forecasting quality is given by

(3.17) $$RM = M_p \;/\; M_x$$

It is called the relative mean square error. It is a 'rate of return' index in that it takes the return to be inversely proportional to the mean square error of forecast, and the cost to be inversely proportional to the mean square error of the benchmark.

If a forecast is 'good', i.e. if it is superior to the benchmark, then a 'natural' scale for this superiority is provided by the relative mean square error in

$$0 < RM < 1.$$

If $RM > 1$, the forecast is prima facie 'inferior'. Normally, the benchmark is provided by an extrapolation of the past series of the target variables. Sometimes 'naive' extrapolations, such as a no change extrapolation ($x_t = x_{t-1}$) or a constant change extrapolation ($x_t - x_{t-1} = x_{t-1} - x_{t-2}$) are used. Complex moving average or autogressive models have also been used. The attraction for extrapolation as a benchmark lies in the fact that it is 'relatively simple, quick and accessible.'[1]

3.3 Extrapolative and Autonomous Components of the Forecast

Having obtained an index for the superiority or otherwise of a model-forecast over an extrapolation, it is worth considering how far the performance of the forecast is due to its use of past history of data regarding the target variable and how far it is due to the exploitation of interrelationships between two or more series. This framework of analysis assumes that extrapolation itself forms a part of all model forecasts. The past history of the variable in question forms a subset of the information used in the model which uses the past history of all variables in the system simultaneously, along with the independently predicted values of exogenous variables.

A forecast P can be thought of as having an extrapolative component P_x due to the use of the past history of the series and a remainder, P_R, due to an analysis of interrelations with other series and the use of future values of exogenous variables. P_R may be called the autonomous component. Thus, we have

(3.18) $$P = P_x + P_R$$

The four concepts being used here are

P = model forecast
X = Extrapolation
P_x = Extrapolative component of P
P_R = Autonomous component of P

---

1. Mincer and Zarnowitz (1969)

What we want to assess is the role of P_x and P_R in generating P.

Since the relative mean square error is defined as $RM = M_p/M_x$, when $RM < 1$, i.e. $M_p < M_x$, the smaller value of M_p can be interpreted as deriving from useful autonomous information.

To judge whether autonomous information has been usefully employed in cases where $M_p \not< M_x$, it is desirable to work with 'linearly corrected' mean square errors. Suppose both P and X are corrected for mean and slope errors such that the corrected mean square errors in the two series are denoted by M_p^c and M_x^c and the respective residuals by U_p and U_x, i.e.

$$(3.19) \qquad M_p^c = M_p - U_p$$

$$(3.20) \qquad M_x^c = M_x - U_x$$

We then have[1]

$$(3.21) \qquad RM = \frac{M_p}{M_x} = \frac{(1 - U_x/M_x)}{(1 - U_p/M_p)} \cdot \frac{M_p^c}{M_x^c} = g.RM^c \text{ (say)}$$

RM^c is the linearly corrected relative mean square error, i.e.

$$(3.22) \qquad RM^c = M_p^c \,/\, M_x^c$$

If X is a best extrapolation, it must be unbiased and efficient. One would expect $M_x^c = M_x$. Since

---

1. We have $M_p^c = M_p(1 - U_p/M_p)$

$M_x^c = M_x(1 - U_x/M_x)$

Dividing, $\dfrac{M_p^c}{M_x^c} = \dfrac{M_p}{M_x} \dfrac{(1 - U_p/M_p)}{(1 - U_x/M_x)}$

or, $\dfrac{M_p}{M_x} = \dfrac{(1 - U_x/M_x)}{(1 - U_p/M_p)} \cdot \dfrac{M_p^c}{M_x^c}$

$$(3.23) \qquad g = \frac{M_p}{M_x} \cdot \frac{M_x^c}{M_p^c} = \frac{M_p}{M_p^c} \quad ,$$

$g \geq 1$ and, therefore, $RM^c \leq RM$ may be correspondingly expected.

In this case when $RM > 1$ but $RM^c < 1$, the forecast may still be relatively more efficient than the extrapolation, implying that autonomous information must have been usefully employed.

The case where $RM > 1$ but $RM^c = 1$ is similar. Here the two mean square errors M_p^c and M_x^c are the same but it does not imply that the two forecasts P^c and X^c (= X) are identical. The implication is that P must have contained autonomous predictive power, and x must have contained predictive power which was not all utilized by P.

Measures of the relative contribution of the autonomous component and of the predictive power contained in the extrapolation that was not utilized by the forecast, can be obtained by relating, in multiple regression, both P and X to the realizations A. The partial correlations of P and A and of X and A must be positive.

The well-known correlation identity

$$(3.24) \qquad (1 - r_{ap}^2)(1 - r_{ax \cdot p}^2) = (1 - r_{ax}^2)(1 - r_{ap \cdot x}^2)$$

provides[1]

$$(3.25) \qquad RM^c = \frac{1 - r_{ap}^2}{1 - r_{ax}^2} = \frac{1 - r_{ap \cdot x}^2}{1 - r_{ax \cdot p}^2}$$

From this relationship it is clear that

1. From equation (3.10), we have, for forecasts corrected for mean and slope errors,

$$M_p^c = (1 - r_{ap}^2)\sigma_a^2$$

$$M_x^c = (1 - r_{ax}^2)\sigma_a^2$$

Dividing, $$RM^c = \frac{1 - r_{ap}^2}{1 - r_{ax}^2}$$

$$(3.26)\quad \begin{cases} RM^c = 1 \text{ when } r^2_{ap\cdot x} = r^2_{ax\cdot p} \\ RM^c > 1 \text{ when } r^2_{ap\cdot x} < r^2_{ax\cdot p}, \text{ and} \\ RM^c < 1 \text{ when } r^2_{ap\cdot x} > r^2_{ax\cdot p} \end{cases}$$

even when $RM^c = 1$, P^c is not identical with X. For that to be so, both partials must be zero.

A value for $r^2_{ap.x} > 0$ means that the forecast P contains predictive power due not only to extrapolation but also to its autonomous component. One can use $r^2_{ap\cdot x}$ as a measure of the net contribution of the autonomous component.

Similarly, $r^2_{ax.p} > 0$ means that X contains predictive power which was not all utilised by P. $r^2_{ax\cdot p}$ can be used as a measure of the extent to which the extrapolative prediction power was not utilised by the forecast P.

From (3.23) and (3.25) we can write

$$(3.27)\qquad RM = g \cdot \frac{1 - r^2_{ap\cdot x}}{1 - r^2_{ax\cdot p}}$$

which indicates that the value of RM or the superiority of P over X depends on g, i.e. the relative mean and slope proportions of the error, and on the relative amount of <u>independent</u> effective information contained in P and X.

Having obtained a measure of the extrapolative prediction power that was not utilised in P, a measure for the extrapolative power that was utilised in the forecast, i.e. an estimate for P_x, can also be designed. We concentrate on P^c rather than P because a good extrapolation is expected to be unbiased and efficient and hence the mean and slope errors are not attributable to it.

The best estimate of P_x is the systematic component of the regression of P on the past values of A.

$$(3.28) \qquad P_t = (\alpha + \beta_1 A_{t-1} + \beta_2 A_{t-2} + \dots + \beta_p A_{t-p}) + \delta_t$$

$$\text{or, } P_t = (\alpha + \sum_{i=1}^{\ell} \beta_i A_{t-i}) + \delta_t$$

The systematic part of (3.28) is an estimate of P_x and the residual δ_t is an estimate of P_R. The coefficient of determination $r^2_{pP_x}$, measures the <u>relative</u> <u>importance</u> of the (autoregressive) extrapolation in generating P^c.

One can simply use $r^2_{P_x}$ as a measure of this relative contribution, but it will be an underestimate of $r^2_{pP_x}$ since P_x is a linear combination of the same variables as X, but the coefficients in P_x are determined by maximising the correlation.

The net contribution of the extrapolative and autonomous components to forecasting performance can be measured by the simple coefficients of determination $r^2_{ap_x}$ and $r^2_{a\delta}$. Further, since

$$(3.29) \qquad r^2_{ap_x} + r^2_{a\delta} = r^2_{ap},$$

the ratios

$$r^2_{ap_x}/r^2_{ap}$$

and

$$\frac{r^2_{a\delta}}{r^2_{ap}} \quad \left(= \frac{1 - r^2_{ap_x}}{r^2_{ap}} \right)$$

can be used to measure the relative contribution of each component to the forecasting power of the adjusted forecast.

3.4 <u>Role of Errors in Exogenous Variables</u>

It was noted in Chapter 1 that the quality of a forecast is necessarily limited by the quality of its input, viz., predictions about exogenous variables. It is, therefore, one of the endeavours in general equilibrium forecasting models to reduce the truly exogenous variables to as few as possible. This implies an attempt to explain most of the variables within the model. In contrast, partial equilibrium forecasting assumes away most of the interrelationships among variables that are in fact operative in the economy. The quality of forecasts generated by these techniques, therefore, becomes highly vulnerable to the quality of predictions regarding the exogenous variables on which the model is dependent. In this context, it is desirable to employ some method by which we can allocate the prediction error as between (i) model misspecification and (ii) errors in the prediction of exogenous variables. The following formulations are based on Theil (1961).

Suppose that a prediction P regarding some variable is generated by a linear single equation model written in the reduced-form:

$$P = b_1x_1^p + b_2x_2^p + \dots b_nx_n^p \tag{3.30}$$

where x_1^p, x_2^p ...x_n^p are predictions regarding the exogenous variables, and b_1, b_2, ... b_n are parameters describing the structure of the model. If there is a constant term, we simply take one of the x's to be equal to 1. For the moment we also ignore the disturbance term.

Equation (3.30) can be rewritten in the form

$$P = BX^p \tag{3.31}$$

where B is a (lxn) row vector

$$(b_1, b_2 \dots b_n)$$

and X^p is a (nxl) column vector

$$\begin{bmatrix} x_1^p \\ x_2^p \\ \vdots \\ x_n^p \end{bmatrix}$$

What we want to find out is how far the prediction error is due to a wrong B (incorrect specification of the model) and how far it is due to a wrong X^p (incorrect prediction of the exogenous variables).

Suppose that the 'true' system is also linear and is given by

$$(3.32) \qquad A = \beta X$$

where A is the 'true' value of the predicted variable, β is a (1xn) vector containing the 'true' values of the coefficients in the system and X is a (nx1) vector of the 'true' values for the exogenous variables.

The error in the predicted variable can now be written as

$$(3.33) \qquad P - A = BX^p - \beta X, \text{ or as}$$

$$(3.34) \qquad P - A = (B - \beta)X + \beta(X^p - X) + (B - \beta)(X^p - X)$$

Given the correct values for exogenous variables, the first term will have a non-zero value depending on the deviation of B from β. This term can, therefore, be said to be due to misspecification of the model. Similarly, given β, i.e. the correct specification of the model, the second term will have non-zero values depending on the deviation of X^p from X. Hence, this term is due to incorrect prediction of the exogenous variables. The third term is due to a combination of both type of errors but is of second order of importance.

It is difficult to imagine, however, that these terms can in practice be calculated. This is so because although we will get to know the true values of the predicted and exogenous variables, the true structure β will rarely be known. In time, however, we can have a larger sample of relevant values and thus obtain an estimate of β which is better than B. One way of estimating the first two terms in equation (3.34) would be to use this new estimate of the coefficients in place of β.

In general, however, suppose that our knowledge of the structure is limited to B. We can rewrite equation (3.34) in the following form:

$$(3.35) \quad \begin{aligned} P - A &= (B - \beta)X + B(X^P - X) \\ &= (BX - A) + B(X^P - X) \\ &= c + d, \text{ say} \end{aligned}$$

Here the last two terms in equation (3.34) have been combined into one term. Every term on the right-hand side can now be available in time. The first term is the same as in equation (3.34). The second term is subject to a different interpretation. It indicates the contribution of error in the prediction of exogenous variables, given the estimated coefficient vector B and not given the true vector β. But this is the best one can do.

If we have n values, for P, A, c and d we can write

$$(3.36) \quad P_t - A_t = c_t + d_t, \text{ or}$$

$$(3.37) \quad \frac{1}{n}\sum(P_t - A_t)^2 = \frac{1}{n}\sum c_t^2 + \frac{1}{n}\sum d_t^2 + \frac{2}{n}\sum c_t d_t$$

Thus, the mean square prediction error can be decomposed into three quadratic terms: the first represents the contribution to error due to model misspecification; the second represents error due to incorrect prediction of exogenous variables; and the third is due to both elements and cannot be allocated between them. Also, in this case, the third term is not of second order of importance.

If there was a disturbance term in the equations, such that the models were:

(3.38) $P = BX^{P} + u^{P}$, and

(3.39) $A = \beta X + u$

where u^{P} and u are respectively the predicted and true disturbance terms, the best procedure then is to substitute the second term in equation (3.34) by

$$\beta(X^{P} - X) + u^{P} - u,$$

and the first term in equation (3.35) by

$$(BX + u^{P} - A)$$

It should be noted that analysis is valid only under the assumption that the 'true' system is of the same type as the one used for prediction.

CHAPTER 4

PREDICTING UNION TAX REVENUES IN INDIA

To test and validate, albeit partially, some of the considerations hitherto outlined, the following Union taxes in India are considered: (i) tax on non-corporate incomes, (ii) tax on corporate incomes, and (iii) import duties. These taxes are of substantial interest in terms of their contribution to Union revenues in India. They also present a forecasting effort with problems typical of a developing economy and as such provide suitable case-studies. To place these taxes in a frame of reference, movements in Union tax revenues in India, are first summarily over-viewed. It is to be noted that only individual tax-revenue models are developed here rather than a tax-interaction model. This is because, in a federal tax structure such as India's, some of the major indirect taxes are levied by the states. A tax-interaction model needs to incorporate tax-rate and base-changes in all individual states apart from those at the centre. A study of this nature is precluded at this stage because of its much wider coverage and requirements.

4.1 Revenues from Union Taxes: An Overview

Import duties and the tax on corporate incomes is levied, collected and retained by the union government in India. Tax on non-corporate incomes, however, is levied and collected by the union government but shared with the state governments. The other major tax to be so shared is Union excise duties. Movements in the relative contributions of these taxes to the Union government revenues are streamlined in Table 4.1.

Table 4.1

Union Tax-Revenues As Percentage of Total Union Tax Revenues: Selected Years

Year	Union's share of Non-Corporate Income-Tax	Union's share of Excise Duties	Corporation Tax	Import Duties	Total of (1), (2), (3) and (4)
	(1)	(2)	(3)	(4)	
1956-57	18.83	34.88	10.37	28.46	92.54
1961-62	08.17	46.69	17.87	22.64	95.37
1966-67	08.87	41.51	17.01	24.78	92.17
1971-72[1]	02.79	56.12	15.93	20.82	95.66

Together, the four taxes account for more than ninety per cent of the union tax-revenues. A wide variety of other taxes, including export duties and taxes on property and capital transactions, thus account for a very small share.

The relative importance of the Union's share of non-corporate income tax is seen to have declined. But this is not due to a fall in revenue from this tax but rather due to an increasing share for the states. In fact, it would be appropriate to predict the non-corporate income tax revenue at the level of collection rather than at the level of the share retained by the Union government. This would permit an abstraction from the arbitrary elements in the distribution of revenues between the centre and the states, and a concentration on the behaviouristic relationships generating the revenues.

The relative contribution of import duties is seen to be steadily losing its position to excise duties. This, of course, is primarily due to increasing import-substitution which decreases the tax-base for import duties and increases that for excise revenues.

1. Percentages for this year are based on revised estimates of revenue data.

For predicting revenue from import duties, its relationship with excise duty revenues would need to be explored.

4.2 Tax on Non-Corporate Incomes

For estimating revenue from this tax, four explanatory variables, viz., per capita income, population and two tax-rate variables, are chosen in line with the reasoning set out in Chapter 2. The influence of per capita income with population held constant was seen to reflect movement of incomes from non-taxable to taxable and from low-rate to high-rate income brackets. Similarly, growth in population with a constant per capita income is taken to account for growth in the tax-base without any change in the distribution of income between brackets.

In the Indian context, it will be necessary to introduce a one-year lag in the income variable because of a distinction between an 'assessment' year and a 'previous' year in the application of tax-rates to incomes. The assessment year is the April 1 - March 31 financial year of the Government of India. The tax-rates for the assessment years are prescribed by a Finance Act every year. But the tax for an assessment year is levied on the income derived in the 'previous' year, which, subject to certain exceptions, is the year ending 31st March immediately preceding the financial year.

To be sure, it may be worthwhile, in a more extensive study, to explore a lag-structure in the estimation of income tax-revenues in greater detail. The actual receipts from the tax in any given year will be different from the tax assessed on the income of that year. The difference arises because some taxes are paid out in advance, and some relate to incomes earned in earlier years. If data about actual tax-receipts and the amount of assessed tax is available for different years, the relationship between these two can be studied by regressing actual tax-receipts on the assessed tax demand. If this is done, assessed tax-demand, rather than actual tax-receipts, may be used as the dependent variable in exploring relationships with income, tax-rates, and other exogenous variables. Predictions about tax-revenues will then be made by two estimated relationships. One will estimate tax-revenue from assessed tax-demand, and the other, tax-demand, from the exogenous variables used in the model.

Alternatively, incomes for the current and some past years may also be used as independent variables along with tax-rates and other exogenous variables in an equation which has actual tax-receipts as the dependent variable. The estimated partial regression coefficients for the current and lagged incomes can be interpreted as relative weights for indicating the contribution of incomes in different years to the tax-receipts of the current year.

In the present context, however, it has not been possible to pursue either of the two alternatives. A lack of sufficient data about tax-demand rules out the first alternative. The second option is not very useful at this stage because of the limited size of the available samples. The introduction of additional income variables will imply a considerable loss of degrees of freedom. However, it must be admitted that in a more extensive study, the lag-structure in income would be worth exploring.

At this stage, we are constrained to use income with a year's lag as the only exogenous income-variable for reasons cited earlier.

Next, independent variables are needed to reflect the influence of tax-rate changes.

The derivation of the tax-rate parameters depicts problems characteristic of developing economies. The search for a rational tax-structure to suit both the growth and income-distribution objectives of these evolving economies, and the practice of introducing ad hoc changes to meet additional revenue requirements lead to a frequent overhauling of the basic tax-structures. In these situations, the derivation of a consistent time series reflecting discretionary changes over a reasonable period becomes a formidable task.

The non-corporate income tax in India has a fairly complicated structure. It is leviable on a wide variety of income-earning entities other than corporations. These categories are individuals, Hindu undivided families, registered and unregistered firms, associations of persons, local authorities and any other non-corporate juridicial person not covered by the above. Different tax-rates prevail for different categories or groups of categories. In addition, there is a fairly detailed scheme of exemptions and rebates, and for this purpose the above categories may have further subdivisions. For example, a distinction is made between an unmarried individual, a married individual with one child, a married individual with more than one child, etc. To catch discretionary changes in the underlying thread of this elaborate structure, a number of simplifying assumptions have to be made.

First, it is hypothesised that tax-rate changes for different categories of assessees are inter-linked and move in the same direction. This is to say that whenever tax-rates are changed for one category, they are accompanied by similar changes for other categories so that their relative positions may not be distorted.

This is generally true except in cases where changes may be brought about for the specific purpose of affecting the relative position of one class vis-a-vis another. Under this assumption, it will be appropriate to concentrate on discretionary changes for one category, say, individuals and to visualise changes for other categories to be reflected by those relating to this category.

Secondly, it is assumed that discretionary changes below a certain limit of income do not have major revenue significance and may be ignored. In the present case, it was decided to concentrate on tax-rates over the income-slabs of Rs. 5000/- so as to abstract from various exemptions allowable below this limit.

Table 4 in the Appendix gives marginal tax-rates over incomes of Rs. 5000/- for an individual for the years 1960-61 through 1971-72. The rates are presented as applicable to specified income-slabs. It is important to have a uniform scheme of income-slabs for the entire sample-period before tax-rate functions can be fitted. This creates a problem when income-margins themselves become revised, as was the case in India in 1962-63, 1964-65, 1965-66 and 1970-71. It was necessary, therefore, to choose a scheme of income-slabs such that the tax-rates applicable in different years can be uniformly written with reference to this.

It will be recalled that in a system with multiple tax-rates it is not feasible to use all the tax-rates as independent variables because of the problem of loss of degrees of freedom and multi-collinearity. Since, in the Indian tax-system, different tax-rates apply to different margins of incomes, it becomes necessary to devise some scheme by which changes in the tax-rates pertaining to different income-slabs may be captured by a limited number of parameters. One such scheme was suggested in Chapter 2. According to this, the tax-rate structure is visualised as a function of two parameters. One parameter stands for, as it were, the 'level' of the tax-rates, and the other for the 'increment' in the tax-rates as we move from one income-slab to the next.

In accordance with this scheme, the following tax-rate functions were fitted for each year of the sample.

(4.1) $r_j = r_y^o + r_y^b \cdot j$ $(j = 1, 2, ...J)$

(4.2) $r_j = r_y^o + r_y^b (m_j)$

where r_j refers to the tax-rate for the jth income-slab. J is the total number of income-slabs, and mj are the mid-points for each income-range. r_y^o's are to be interpreted as tax-rate 'levels' and r_y^b's as the 'incremental' factors.

The results are reported in Table 6 and 7 of the Appendix. Judged from the t-values of the regression coefficients and the correlation between the dependent and the independent variables, both the fits are found to be good. The next step was to try the two sets of estimated r_y^o's and r_y^b's, i.e. the income-tax rate levels and the incremental factors as independent variables in an equation explaining non-corporate income-tax revenue. The other independent variables used were per capita income with one-year's lag and population in consonance with the reasoning set out earlier. Because some of the values of r_y^o in the first set of the tax-parameters are negative, log-linear form of the equation was not used in this case. For the second set, both linear and the log-linear forms were tried.

Table 4.2 gives the estimated coefficients and related statistics. On a priori reasoning, it would be expected that the partial regression coefficient for per capita income would have a

Table 4.2

Regression of Revenue from Taxes on Income other than Corporation Tax on Given Variables: 1961-62 to 1972-73

Form	R-SQ (i) unadjusted (ii) adjusted	D-W Statistic	Intercept	Independent Variables: Per Capita Income with one year's lag	Population	Tax-rate Parameters (first set) r^o_y	r^b_y
Non-log	0.99 0.98	1.95	-1832.0 (-4.0)	-0.26666 (-0.6)	4.0782 (2.8)	500.39 (1.1)	5052.9 (2.0)
						Tax-rate Parameters (second set)	
log	0.99 0.99	1.83	-27.788 (-3.1)	-0.1415 (-0.3)	5.4816 (3.0)	0.76935 (2.4)	-0.22317 (-1.4)
Non-log	0.99 0.98	1.97	-1842.5 (-4.0)	-0.2846 (-0.6)	4.1236 (2.8)	491.00 (1.2)	44068.5 (1.7)

Figures in brackets are t-ratios.

positive sign, i.e. income-tax-revenues would go up with an increase in per capita income when the other variables are held constant. This presumes that an upward shift in per capita income moves people from non-taxable to taxable, and from low-rate to high-rate income brackets. The signs obtained for the regression coefficients for per capita income are, however, negative in all the equations in Table 4.2. But not much importance can be attached to these negative coefficients because of the existence of multicollinearity among the regressors. The coefficient of correlation between per capita income and population is 0.99.

The expected signs for the regression coefficients of population and two tax-rate parameters are positive. Except for r_y^b in the second equation in Table 4.2, positive signs are obtained as expected. The significance of these individual coefficients can be judged in a one-tail test. The hypothesis to be tested is whether an individual coefficient is significantly greater than zero. The critical value of t at a 5% level of significance for 7 degrees of freedom is 1.895. The coefficient obtained for population in all the fits is observed to be significantly positive.

The results are not so unambiguously encouraging for the tax-rate parameters. The coefficient of r_y^o in the first equation is not significantly greater than zero. The coefficient of r_y^b in the second equation is negative. But the tax-rate coefficients in other cases bear well against a critical value of $t = 1.895$.

In view of the inadequacy of per capita income as an appropriate regressor in this sample, and the high correlation with population, it was decided to lump these two variables into one. As a consequence, gross domestic product at current prices was used as an explanatory variable in the place of the above two.

Table 4.3

Regression of Revenue from Taxes on Income Other than Corporation Tax on Given Variables: 1961-62 to 1972-73

Form	R-SQ (i) Unadjusted (ii) Adjusted	D-W Statistic	Intercept	GDP at current prices with 1 year lag	Tax-rate parameters (first set) r_y^o	r_y^b
Non-log	0.98 0.97	1.71	-376.06 (-3.2)	0.0143 (15.7)	1107.36 (2.8)	6897.11 (2.9)
					Tax-rate parameters (second set)	
log	0.99 0.98	1.39	-3.54 (-2.5)	1.044 (19.7)	1.289 (4.7)	-0.0626 (-0.3)
Non-log	0.98 0.97	1.68	-384.65 (-3.3)	0.0147 (15.5)	1041.85 (2.9)	21079.2 (0.7)

(Figures in brackets are t-ratios)

Again, equations were estimated with the two sets of tax-rate parameters. These are given in Table 4.3.

All the equations have a high coefficient of determination (R^2). The significance of individual coefficients sheds more light on the choice of predictor variables and functional forms. Again, the expected signs for the coefficients of the three independent variables are positive. This implies that an increase in G.D.P. or in any of the tax-rate parameters will lead to an increase in the non-corporate income tax revenue. A one-tail test can be used to judge whether the estimated coefficients are significantly positive. The critical value of t for a 5 per cent significance level and 8 degrees of freedom is 1.860. In the linear fit, with the first set of tax-rate parameters, all the regression coefficients are significantly positive.

When the second set of tax-rate parameters are used, the coefficients for the tax-rate incremental factor (r_y^b) become insignificant in both the linear and log-linear fits. In the latter case, the coefficient also has the wrong sign. The coefficients for the other two variables, however, are significant and have the expected signs. Judged from the t-values for the coefficients of these variables, the logarithmic form seems to be better.

Because of this reason, a logarithmic form was also tried with the first set of tax-rate parameters. Since some of the values of r_y^o are negative in this set, it was possible to fit only a partially logarithmic form of the following type.

$$(4.3) \qquad \log R_y = a + b_1 \log(G.D.P.) + b_2 r_y^o + b_3 \log(r_y^b)$$

The results obtained for this equation seem to be the most promising. All the coefficients are now significantly positive and the t-values for these have improved over the earlier fits. The estimated equation and other relevant statistics are given as follows.[1]

$$(4.4) \qquad \log Ry = \underset{(-0.88)}{-0.937} + \underset{(20.3)}{1.0075}\log(G.D.P.) + \underset{(4.9)}{4.6148}(r_y^o) + \underset{(4.1)}{1.1507}\log(r_y^b)$$

R-SQ: Unadjusted 0.99
Adjusted 0.98
D-W: 1.34

In view of the level of significance of the estimated coefficients, this equation seems to be appropriate for predicting non-corporate income-tax revenues. The value of t at the level of significance of .005 for 8 degrees of freedom is 3.355 in a one-tail test. All the estimated regression coefficients in (4.4) are significant at this level. Alternatively, the first equation in Table

1. Figures in brackets are t-ratios.

4.3 may be used. The coefficients of regression in this equation also are significantly positive at the 5 per cent level for all the variables. Further, this equation is of a simple linear form. Equation (4.4) above is a hybrid one and its constant term does not seem to be significant. We shall use equation 1 in Table 4.3 for prediction purposes.

4.3 Tax on Corporate Incomes

For corporation tax revenue, again a one-year lag is introduced in the income variable of the regression equation because of the distinction between a 'previous' year and an 'assessment' year. This is in line with the reasons outlined for the income tax on non-corporate assessees.

In fact, here again the lag-structure in the income-variable is worth exploring in greater detail. But for the reasons pointed out in the previous section, we restrict ourselves to introducing just a one-year-lag in the income variable. For a more extensive study, the lag-structure in income would be a direction worth pursuing.

In addition to income, other exogenous variables are needed to reflect discretionary changes affecting the revenue from corporation taxes.

The derivation of a consistent series for reflecting movements in tax-rates and bases is, however, again full of complexities typical of a developing economy. Within the reference period of 1961-62 through 1972-73, corporate incomes have intermittently been subject to a variety of taxes such as (i) income-tax, (ii) super-tax, (iii) super-profits tax, (iv) surtax, (v) tax on inter-corporate dividends, (vi) tax on bonus issues, (vii) tax on excess dividends; and (viii) surcharges on some of these taxes. The resultant structure closely meets Kaldor's (1956) earlier description of it as '... a perfect maze of unnecessary complications ....' At the very least, it depicts the typical trial-and-error process of a developing economy in finding out suitable alternatives to its revenue requirements in conjunction with its desired resource allocation and income-distribution objectives.

It is clear that the tax-rates pertaining to all the above taxes cannot be used as separate regressors because of the implied loss in degrees of freedom in a sample of limited size. Since these taxes are applicable, in general, to different tax-bases, it is also not possible to collapse them into one. A number of simplifying assumptions have, therefore, to be made. First, only income tax and super tax, which were, in fact, integrated into one in 1965-66, are covered in conjunction with the (i) super profits tax which was operative only in 1963-64 and the (ii) surtax which later replaced it. The income-tax and super-tax, in fact, account for the major share of revenues from taxes on corporate incomes. Discretionary changes in other taxes are abstracted from on grounds that either these are reflected in the changes relating to the taxes which are included, or that they were of minor revenue significance. Their main purpose was the restriction of some of the undesirable practices of the corporations rather than revenue generation. Further, income-tax and super-tax-rates for only non-dividend incomes are covered.

Table 5 in the Appendix gives the tax-rates for non-dividend incomes for the period 1961-62 through 1972-73. The rates are the sum of income-tax and super-tax rates till 1964-65. Later the two taxes were integrated. The tax-rates differ with the type of company, the type of income and the level of income. The classifications, however, have changed over time and for building a consistent time-series of tax-rates, a homogenous classification for all the years has to be chosen. For this purpose the classification introduced in 1966-67 which prevails to date seemed a suitable choice. Tax-rates for earlier years were written in accordance with this scheme.[1]

For corporate income tax, however, tax-rate functions cannot be fitted as was done in the case of non-corporate income tax. The tax-rates in the latter case related to income-slabs arranged in ascending order. Hence, it was possible to interpret the tax-rate parameters as providing the 'level' of taxation and the 'degree of progression'. Further, it was possible to introduce other variables such that the effect of movement of incomes from lower income-slabs to higher income-slabs could be studied. This cannot be duplicated for the corporate income tax because the classification according to which tax-rates are arranged is not one of ascending incomes but rather one of type, size and the source of income of the corporations.

It is also not possible to use the tax-rates pertaining to different categories of companies and/or incomes as separate regressors because they would exhaust all the degrees of freedom. One suitable alternative in this situation is to build a series of weighted averages of tax-rates where the weights are derived by the relative contribution of each category of company

---

1. This involved a degree of arbitrariness for two years, viz., 1964-65 and 1965-66 in which a different and much more detailed system of classification was prevalent. For earlier years, the prevailing classification could simply be extended into the one which is used. The nature of adjustments involved are given in the notes to Table 5 in the Appendix.

or income to corporate tax-revenue. This procedure would, however, be feasible only when extensive data about the assessed corporate incomes in different categories is available.

In the absence of such data, which is presently the case, other possibilities may be explored. It was suggested in Chapter 2, that in situations like these, 'principal components' of the tax-rates may be computed. The first few components can then be used as independent variables to reflect discretionary changes in tax-rates.

In the present case, the first two principal components were used along with other exogenous variables in a regression equation with revenue from corporation tax as the dependent variable. Together the two components explain about 85 per cent of the variation in the tax-rates. The introduction of more principal components in the regression equation would have accounted for a somewhat greater variation in the tax-rates. But in doing this more degrees of freedom would have been lost. In view of the small size of the sample, greater importance was given to the latter consideration. The first two components are given in Table 2 of the Appendix. The other independent variables were the following: (i) G.D.P. at current prices with one year's lag, (ii) super profits/surtax rates. G.D.P. at current prices was used as a proxy for company incomes. Since tax-rate functions were not estimated for this tax, for reasons cited earlier, it was not required to study the influence of growth in income through per capita income and population so as to reflect effects of distribution of income between companies. Super profits/ surtax rates provide a consistent series[1] for all types of companies

1. Except for the year 1963-64 when the base of the tax was somewhat different to that for the later years. See Notes to Table in the Appendix.

and incomes, and hence it was possible to use this as a separate regressor. The influence of basic income tax and super tax rates was reflected by the principal components.

Since increments in G.D.P. or superprofits/surtax rates would augment revenue from the corporation tax, the signs of the coefficients of these two variables are expected to be positive. It is difficult to be categorical about the expected signs of the coefficients of the principal components. The principal components are 'artificial' variables and unless an 'economic' interpretation can be attached to them, the nature of the relationship between these and the dependent variable cannot be defined a priori. In the present case, some values of the principal components are negative and it is difficult to assign any meaning to them with reference to tax-rates.

The results of the fit are given in Table 4.4. In this case only the linear form of the equation was tried. Because of zero and negative values in the data matrix, a transformation into natural logarithms was not possible. The linear form implies that the partial flexibility coefficients ($\partial y/\partial x_1$, etc.) are constant.

Table 4.4

Regression of Revenue from Corporation Tax on Given Variables: 1961-62 to 1972-73

Form	R-SQ (i) Unadjusted (ii) Adjusted	D-W	Intercept	Independent Variables: GDP at Current prices with 1 year's lag	Superprofits/ Surtax rates	Principal Component 1	Principal Component 2
Non-log	0.85 0.76	1.91	-266.94 (-1.49)	0.0183 (3.79)	562.21 (1.97)	-45.426 (-1.85)	66.007 (1.80)

Figures in brackets are t-ratios.

For one-tail tests, the critical value of t at the 5% significance level for 7 degrees of freedom is ± 1.895. The coefficients of G.D.P. and superprofits/surtax rates are seen to be significantly greater than zero. Since the expected signs for the coefficients of the principal components are ambiguous, the above critical value can be used for a two-tail test at the 10 per cent significance level. The estimated t-values for the coefficients of the principal components marginally compare with this critical value. Thus, it is only at the 10 per cent significance level that the coefficients can be said to be significantly different from zero.

To gain more idea of the expected signs of the principal components, correlations of these with individual tax-rate series given in Table 5 of the Appendix were studied. The second principal component is highly correlated with the first set of the tax-rates, i.e. the tax-rates applicable to domestic companies in which the public are substantially interested, and which have an income less than Rs. 25,000. The coefficient of correlation between these two is 0.965. The first principal component was found to be highly correlated with the tax-rate series of all other types of companies/incomes except the first. In some sense, the two principal components can be visualised as reflecting movements in the tax-rates with which they are highly correlated.

We would expect a positive sign for the second principal component if it can be assumed that increments in the tax-rates applicable to domestic companies with incomes less than Rs. 25,000 and in which the public are substantially interested would lead to a rise in the tax-revenue. This is reasonable to expect. This means that a one-tail test could have been applied on the coefficient of the second principal component. The estimated t-value is then nearly equal to the critical value of t at the 5% significance level.

The sign of the coefficient of the first principal component is negative. This component, on the strength of the observed correlations, is seen to reflect movements in tax-rates applicable to categories of companies other than the first among those given in Table 5 of the Appendix. These categories consist mainly of companies with a wide capital base. A negative sign for the estimated coefficient means that an increment in the tax-rates applicable to incomes of these categories of companies leads to a fall in the tax-revenue. A tentative hypothesis to explain this may be that as tax-rates increase, these companies have a tendency to invest more in capital goods and thus reduce the tax-base.

4.4 Import Duties

Economic development has a twofold effect on the tax-base for import duties, i.e. taxable imports. First, there is an income effect. As domestic incomes go up, demand for imports also goes up. Second, there is a substitution effect. This may be due to a fall in the price of domestic goods relative to the price of imported goods, or this may be due to more extraneous reasons. With economic development fairly underway, most of the developing countries undergo bouts of import substitution. This may be effected by quantitative restrictions which override the opposite price incentives whenever imported goods are cheaper relative to domestic goods.

In India, the effect of import-substitution on the revenue from import duties is visible in the decline of its relative share in total Union tax-revenues. This can be observed from Table 4.1.

Correspondingly, the share of Union excise duties has gone up because its tax-base, i.e. domestic production, has increased with import substitution. The link between the two tax-revenues is important. Growth in the tax-base for excise duties undermines the tax-base for import duties. This has a negative effect on the revenue from import duties. On the other hand, there is also a positive effect. This is due to a system of 'countervailing duties' which prevails in India.

The countervailing duty is an additional duty in the Indian Tariff Act. Because of this all imported goods have to pay, in addition to the basic customs duty, an additional amount at the rate of the central excise duty, if it is leviable on similar goods manufactured within the country. The purpose of the countervailing duties is obviously protective. But its contribution in augmenting revenue from import duties has been significant. In fact, it represents the major avenue whereby discretionary changes in the tax-rate and base for import duties are brought about. The general rates of import duties are fairly rigid because of bilateral and international economic agreements about tariffs and trade.

To reflect discretionary changes brought about by the additional countervailing duties each year, it was decided to use the first differences in the revenue from Union excises. The difference between any year's excise revenue, and its preceding year, is due to two reasons: (i) automatic growth in revenue, and (ii) growth in revenue due to discretionary changes in the current year. It is the latter component which directly affects revenue from import duties. The use of first differences in the automatic growth would, therefore, be biased to the extent of the first component. The appropriate alternative, to be sure, is to build one or more series of tax-rate parameters, such as the ones used for non-corporate income tax, and use these as the independent regressors. However, the difficulties in building up a series to reflect the tax-rate changes and the effects of countervailing duties seem to be prohibitive at this stage. Some of the import duties in India are specific, and some, ad valorem. Some of the duties have changed from specific to ad valorem within the stipulated sample period. There is a wide variety of standard and preferential rates and a very extensive classification of goods. The system of classification for excise duties is even more extensive and complex. This renders the building of a series of basic import duties with countervailing excise duties superimposed on them even more difficult. The problem is further confounded because of an element of arbitrariness in the imposition of excise duties. This arbitrariness is due to a system of 'notification' which allows the assessing officers powers to change the excise duties originally proposed for any assessment year and incorporated in the Statutes. These difficulties preclude the estimation of rate-of-duty functions and the subsequent use of tax-rate parameters in revenue estimation equations in the present context.

As noted earlier, this should not prove to be too limiting inasmuch as the basic import duties are fairly rigid and the effect of countervailing duties can be reflected by using the first differences of revenue from excise duties as an independent variable. The relationship between this series and the revenue from import duties is expected to be positive because it reflects increases in the effective rates of import duties.

In the estimation of a prediction equation for revenue from import duties, account has also to be taken of the devaluation of the Indian rupee in 1966. This has not only behaviouristically changed the demand for imports but has also rendered import data before 1966-67 incomparable with those for the latter years. The former are written in terms of the pre-devaluation rupee and the latter, in terms of the devalued rupee. For these reasons, it is necessary to use a dummy variable which takes a value 1, say, for years 1966-67 and onwards, and a value zero for the earlier years.

First, a multiple regression of the revenue from import duties on gross domestic product at current prices , the dummy variable, the first differences of revenue from excise duties, and the ratio of unit value of imports to domestic wholesale prices was tried. The last variable was included to account for changes in imports due to the relative movement of import prices and domestic prices. It is expected that when imported goods become relatively costlier, the demand (quantity) for imports would fall, and consequently the revenue from import duties would also decline. Hence, a negative relationship between import duty revenue and the index of unit value of imports upon domestic prices is expected. On the other hand, because of the increase in import prices, the <u>value</u> of imports will go up and there will be a positive effect. The sign of the net effect will depend on the relative strength of the two effects.

The estimated equations with import duty revenue as the dependent variable are given in Table 4.5.

The significance of individual coefficients is to be tested according to the expected signs. Where the expected signs are unambiguous, one-tail tests may be applied to test whether the coefficient is significantly greater or less than zero as the case may be. Where the expected sign is ambiguous, a two-tail test is needed. This will test whether the coefficient is significantly

Table 4.5

Regression of Revenue from Import Duties on Given Variables:[1] 1960-61 to 1970-71

Form	R-SQ (i) Unadjusted (ii) Adjusted	D-W	Intercept	Independent Variables			
				Dummy Variables	GDP at Current Prices	Unit Value of Imports upon Domestic Wholesale Prices	First Differences Of Excise Duty Revenues
Non-log	0.74 0.57	1.04	409.98 (2.8)	-760.32 (-2.97)	0.0148 (1.59)	-23.25 (-2.8)	-0.8555 (-0.98)
Log	0.76 0.59	0.71	11.55 (0.95)	-1.957 (-2.45)	1.3515 (2.3)	-4.095 (-1.9)	-0.0670 (-0.22)

Figures in brackets are t-ratios

different from zero.

For one-tail tests, the critical value of t at a 5% level of significance is ± 1.945 for 6 degrees of freedom. For a two-tail test, the corresponding critical value of t is ± 2.447.

In the linear fit, the coefficient of G.D.P. is observed to be not significantly greater than zero. But it is so in the log-linear fit. On the other hand, the coefficient of unit value of imports upon domestic wholesale prices is significantly different from zero in the linear fit, but not so in the log-linear fit. The signs obtained for the coefficient of the first difference of excise duty revenues are contrary to expectations in both the fits. But this may be due to multicollinearity. The coefficient of correlation between this variables and G.D.P. at current prices is 0.86.

To test the significance of the coefficient of the dummy variable, again a two-tail test is needed. It should be noted that the dummy variable stands here for the <u>difference</u> between the pre- and post-devaluation intercepts. Devaluation has a positive effect

1. It should be noted that for import duties, the sample period is only of 11 years as compared to 12 for the earlier two taxes. This is because data for 1971-72 for the index of unit value of imports was not yet available.

on import revenues because it increases the value of imports in terms of the domestic currency, and a negative effect, because it reduces the quantity of imports. Hence, the sign of the coefficient of the dummy variable can go either way. Judged against a critical value of -2.447, the estimated coefficients are significantly different from zero in both the fits. Although it is difficult to carry out significance tests on the D-W statistic because of the small size of the sample, the low values of this statistic in the present case should be noted. These may simply mean that the accuracy of the estimated coefficients is overstated and hence care is needed in interpreting these coefficients.

Thus, overall, the equations do not seem to be very satisfactory. There is a problem of multicollinearity which may be the reason why wrong signs are being obtained for the excise revenue variable. In addition, in the linear fit, the coefficient of G.D.P. is not significantly positive; and in the log-linear fit, the coefficient of the relative price variable is not significantly different from zero. Thus, neither of the two equations in Table 4.5 seem appropriate for predicting import duty revenues.

It was, therefore, decided to break the estimation procedure into two parts so that the imports and the import duty revenues may be estimated separately. This may save some degrees of freedom, enable us to get round the problem of multicollinearity, and highlight the problematic part of the estimation procedure when only a single equation is used.

Revenue from import duty was now estimated from three variables, viz., a dummy variable to take account of the 1966 devaluation, first differences in excise duty revenues to take account of the discretionary changes brought about by countervailing duties, and imports in money terms. The results obtained for the revenue function are given in Table 4.6.

Table 4.6

Regression of Revenue from Import Duties on Given Variables, 1960-61 to 1970-71

Form	R-SQ (i) unadjusted (ii) adjusted	D-W	Intercept	Independent Variables		
				Dummy Variable	Imports	First Differences of Union Excise Revenues
Non-log	0.79 0.70	1.98	-772.82 (-3.2)	-539.15 (-3.7)	0.722 (4.6)	2.2506 (3.1)
log	0.86 0.80	1.89	-18.89 (4.4)	-1.355 (-4.4)	2.998 (5.5)	0.7269 (3.7)

Figures in brackets are t-ratios

The significance of individual coefficients is again tested in a one-tail or two-tail test according to the expected signs of the coefficients. The expected relationship between money imports and import revenue, and that between the first difference in excise duty revenue and import revenue is expected to be positive. The critical value of t for these two cases is 1.895. It provides a 5 per cent level of significance for 7 degrees of freedom. The coefficients obtained for these variables are observed to be significantly greater than zero in both the equations in Table 4.6. The values of the D-W statistic are also much better here than those reported in Table 4.5.

For the dummy variable, a two-tail test is needed. The critical value of t in this case is 2.365 for a 5 per cent significance level. The coefficient of the dummy variable is also seen to be significant in both the equations.

On the basis of the improvement in the t-values for individual coefficients, the logarithmic form of the equation may be preferred to the non-logarithmic form. The improvement in the value of R^2, both adjusted and unadjusted, may also be noted. This in itself cannot, however, be used for choosing one form over the other.

Having obtained a revenue function, the second step was to develop an equation for demand for imports. If this could be obtained, predictions can be generated by a reduced-form equation which is derived by substituting the import-equation into the revenue function.

However, the estimation of demand for imports does not seem to be very promising in the present context. The behaviouristic effects of income and price movements on the demand for imports is generally distorted in a developing economy because of various quantitative restrictions. Furthermore, in a country like India, there are substantial ad hoc changes in imports because of recurring food crises and other emergent situations. Barring these difficulties, the demand for imports can be seen as a function of real income and the price of imports relative to domestic goods.

As a first attempt, two variables, viz., gross domestic product at constant 1960-61 prices and an index of unit value of imports upon domestic wholesale prices were used as independent variables. Real imports (1960-61=100) were used as the dependent variable. This series was obtained by deflating money imports by an index of unit value of imports. In addition, a dummy variable was used as an independent variable to distinguish between pre- and post--devaluation intercepts.

The expected relationship between demand for imports and real income is expected to be positive for all non-inferior imports. On the other hand, the relative price of imports is expected to have a negative relationship with the demand for imports.

The estimated relationships are given in Table 4.7. The critical value of t for 7 degrees of freedom is ± 1.895 for a 5% significance level. This is to test whether a coefficient is significantly greater than or less than zero, as the case may be. None of the estimated coefficients in the equations in Table 4.7 are significant. In fact, the fits obtained are so bad as to have a negative value for the adjusted R^2.

Table 4.7

Regression of Real Imports on Given Variables: 1960-61 to 1970-71

Form	R-SQ (i) Unadjusted (ii) Adjusted	D-W	Intercept	Independent Variables		
				Dummy	GDP at 1960-61 Prices	Index of Unit Value of Imports Upon Domestic Wholesale Prices (1960-61=100)
Non-log	0.067 -0.332	1.00	166.64 (1.28)	-5.936 (-0.15)	-0.00059 (-0.51)	-0.5275 (-0.42)
log	0.012 -0.398	0.75	4.84 (0.88)	-0.0408 (-0.11)	0.0347 (0.15)	-0.1144 (-0.12)

Figures in brackets are t-ratios

If more satisfactory results had been obtained for the demand for imports in quantity terms, the procedure would have been to multiply the entire equation by import prices before substituting it in the revenue equation. This is because, in the revenue equation, money imports rather than real imports are treated as an exogenous variable.

For this reason, an attempt was made to predict the demand for money imports directly. The same set of the independent variables were used except that G.D.P. at current prices was also tried. On a priori grounds, better fits are expected because import prices enter both sides of the equation.

In this case, the expected relationship between the relative price of imports vs. domestic goods and the money imports is not necessarily negative. As the import prices relative to domestic prices rise, the real demand for imports is expected to fall. Thus, there is a negative 'quantity' effect. But since this lower quantity will be multiplied by the increased import prices before the value of imports is obtained, there is a positive 'price' effect. The net relationship, therefore, depends on the relative strength of the two effects.

The relationship between G.D.P. and money imports can also be said to be ambiguous. In general the expected relationship is positive. However, with economic development, of which G.D.P. is an index, a substantial degree of import substitution takes place, and this may affect the positive relationship. In some sense, imports become 'inferior' goods.

Table 4.8 gives the relevant statistics for alternative fits in an attempt to estimate the demand for money imports. The first two equations incorporate G.D.P. at constant prices as one of the independent variables. In the latter two equations, G.D.P. at current prices was used.

Table 4.8

Regression of Money Imports on Given Variables: 1960-61 to 1970-71

Form	R-SQ (i) Unadjusted (ii) Adjusted	D-W	Intercept	Independent Variables: Dummy	GDP at Constant Prices	Index of Unit Value of Imports/Domestic Wholesale Prices (1960-61 = 100)
Non-log	0.85 0.78	1.03	4179.88 (2.3)	22.63 (0.04)	-0.01915 (-1.16)	-26.899 (-1.5)
log	0.83 0.76	0.75	11.879 (2.2)	-0.0408 (-0.11)	0.03467 (0.15)	-1.114 (-1.1)
Non-log	0.85 0.78	1.04	4198.88 (2.3)	22.746 (0.04)	-0.01839 (-1.18)	-27.039 (-1.5)
log	0.83 0.76	0.75	11.945 (2.2)	-0.0399 (-0.11)	0.0299 (0.13)	-1.119 (-1.1)

Figures in brackets are t-ratios.

However, the coefficients of individual variables do not seem to be very promising in any of the equations. Because of the ambiguity of expected signs in the case of each of the independent variables, a two-tail test is needed in all the cases. The critical value of t at a 5 per cent significance level for 7 degrees of freedom is $\pm$ 2.365. In comparison to this value none of the coefficients obtained is significantly different from zero in any of the fits. This is so also for the coefficients of the dummy variable. The equations were reestimated after dropping the dummy variables. But the fits do not seem to improve. Furthermore, the low values of the D-W statistic raise doubts as to the accuracy of the estimated coefficients. It is clear that none of these equations can satisfactorily be used to estimate the demand for imports.

The reasons for the distortion of a demand for import function in a country like India presumably lie, as has already been noted, in the various quantitative restrictions on imports in the form of quotas, etc., and in the component of imports which is due to ad hoc reasons such as droughts and wars.

It does not seem worthwhile, therefore, to continue the attempt to generate an income-based forecast of import duty revenues at this stage. Rather it seems best to take imports themselves as the exogenous variable and use equation 2 in Table 4.6 as the prediction equation.

4.5 Concluding Observations

On the basis of the considerations given in the previous sections, the following equations may finally be presented as useful for prediction purposes.

For the income-tax on non-corporate assessees, the estimated model is given by the following:

(4.5) $$R_{y,t} = -376.06 + 0.0143\ Y_{t-1} + 1107.36\ r^{o}_{y,t} + 6897.11\ r^{b}_{y,t}$$

where $R_{y,t}$ = non-corporate income-tax revenue in year t
Y_{t-1} = G.D.P. at current prices in year t-1
$r^{o}_{y,t}\ r^{b}_{y,t}$ = tax-rate parameters for year t calculated with reference to equation (4.1).

For corporation tax, the following estimated equation is suggested.

(4.6) $$R_{c,t} = -266.94 + 0.0183Y_{t-1} + 562.21\ r_s - 45.426\ PC_1 + 66.007\ PC_2$$

where $R_{c,t}$ = revenue from corporation tax in year t
Y_{t-1} = G.D.P. at current prices in year t-1
r_s = superprofits/sur tax rates
PC_1 = first principal component of corporation tax rates
PC_2 = second principal component of corporation tax rates.

For the revenue from import-duties, the following equation may be used.

(4.7) $$\log R_{m,t} = -18.89 - 1.355X + 2.998 \log M_t + 0.7269 \log \Delta R_{e,t}$$

where $R_{m,t}$ = revenue from import duties in year t.
$X = \begin{cases} 0 \text{ for pre-1966-67 years, and} \\ 1 \text{ for 1966-67 and later years} \end{cases}$
M_t = money imports in year t
$\Delta R_{e,t}$ = increment in revenue from Union excise duties in year t over last year.

Based on the above equations, the following elasticities are estimated. These give the percentage change in the tax-revenues following a 1 per cent change in given exogenous variables. Since the equations for the non-corporate income tax and the corporation tax are linear, elasticities are calculated with reference to the mean points. For the principal components the elasticities were calculated at last year's values rather than the mean points.

Table 4.9

Estimates of Partial Elasticities

Revenue	Exogenous Variables							
	Y_{t-1}	$r^o_{y,t}$	$r^b_{y,t}$	r_s	$PC_{1,t}$	$PC_{2,t}$	M_t	$\Delta R_{e,t}$
R_y	1.0147	.0084	1.046					
R_c	1.383			0.426	-0.1494	0.0043		
R_m							2.998	0.727

In conclusion, the limitations of the above models may be noted. The models which we have developed are all single-equation models. It has not been possible to adopt the whole scheme which was suggested in Chapter 2. In particular, interaction among taxes could not have been allowed such that different tax-revenues could be simultaneously determined. In a federal tax-structure such as India's, a tax-interaction model would have to take account of major state taxes in addition to the union taxes. This would involve a study of discretionary changes in tax-rates and bases in all states. This necessitates a much bigger forecasting exercise than the one which was stipulated here.

What has been done here is still worthwhile in that, within the field of partial equilibrium models, an attempt has been made to take account of tax-rate changes in a system with multiple tax-rates. It has been noted before that the methods by which this has been done in earlier literature on the subject are not satisfactory. Three different methods were employed for this purpose in the three taxes which were considered here. In our view, these methods provide more useful answers.

The basic method which was suggested in Chapter 2 for taking account of discretionary changes in tax-rates was that of deriving 'tax-rate functions'. This method could not be applied uniformly to all taxes in the present case-study. In fact, the difficulties which are likely to arise whenever an empirical study is considered were anticipated. The complexities and individual characteristics of the fiscal sector of the developing economies, and the lack of appropriate data, necessitates adjustments in the techniques of formulation and estimation of revenue-forecasting models. The variety of the cases that fall within the category of developing economies ensures that we cannot hope to have a uniformly applicable revenue forecasting model. All that it is possible to have is a basic core of a set of techniques in which additional adjustments are needed in the case of each empirical study.

CHAPTER 5

MODEL- AND BUDGET-ESTIMATES: A COMPARATIVE EVALUATION

A prediction model or method is appropriately evaluated when, inter alia, predictions for periods beyond the sample are compared with corresponding actual values. This, of course, is only possible at a future date, i.e. when such actual values become available. However, some idea of the quality of a forecasting model is obtained by estimating the 'within-sample' values and comparing these with the corresponding actual values.

The basic concepts in the evaluation of forecasts were outlined in Chapter 3. It will be recalled that a study of the deviation of predictions from realizations provides useful information about the size and nature of prediction errors. Prediction errors are analysed with the help of concepts such as the mean square error, the inequality coefficients, and the inequality proportions. The study of these statistics falls within the purview of 'absolute accuracy analysis'. In addition, however, it is important to know how one prediction method or model performs in comparison to another. This is the subject of 'relative accuracy analysis'. For an analysis of this nature,competing forecasts with similar sample periods and objective variables are needed.

An attractive set of tax-revenue estimates for the purpose of comparison is the budgetary estimates of the Indian Ministry of Finance. The budget, or technically, the 'annual financial statement' of the Government of India, contains the estimates of the expected tax-receipts for the current financial years. These estimates account for both the automatic growth in tax-revenues and the effects of the budgetary proposals in the current year. There is no explicit forecasting model for the preparation of these estimates. In general, it may be assumed that they are based on the opinions of experts and on historical information about movements in tax-revenues, especially in response to previous discretionary actions. An analysis of the relative accuracy of budget estimates would give an idea of the fiscal marksmanship of the Ministry of Finance in India. Here, we consider the budgetary predictions for the three Union taxes considered in Chapter 4.

Both the model-estimates and the budget-estimates are compared to a set of 'benchmark' predictions. In the literature on evaluation of forecasts, such predictions have usually been obtained by 'naive' models such as a 'no-change' or a 'constant-change' extrapolation. In certain cases, more sophisticated autoregressive or moving-average schemes have been used. We have chosen a simple autoregressive model where the current value of a tax-revenue is seen simply as a function of its value in the previous year. An equation has been estimated for each of the three taxes mentioned above for sample periods similar to those used in the previous chapter.[1] Revenues estimates are then generated from these equations for each year in the sample. This set of estimates will now be referred to as 'extrapolation'. For purposes of comparison with actual data, we now have three sets of predictions for each tax, viz., model-predictions, budget-estimates, and extrapolations.

It must be conceded that the appropriate framework for an

1. Since one-period lag models are being considered, one extra value of the lagged independent variable was needed for the year previous to the first year of the samples.

evaluation of the model forecasts is a comparison of 'beyond-sample' predicted values with their corresponding realizations. This, as has already been admitted, is not presently possible. An alternative is to re-estimate the models with a smaller sample and to compare the predictions for the remaining few years in the sample with the available actual values. However, the very limited size of the available samples constrains us from pursuing this alternative.

Before individual taxes are considered, it is as well to recapitulate some of the parameters which are to be used in the absolute accuracy analysis, previously described and analysed in Chapter 3. The symbol P_t stands for an individual predicted value for period t, and A_t for the corresponding realization. $\bar{P}$ and $\bar{A}$ are, respectively, the arithmetic means of the predicted and the realized series, and n is the size of the sample. Some of the statistics to be used in this chapter are defined as follows.

(i) Bias: $(\bar{P} - \bar{A})$

(ii) Mean Square Prediction Error:

$$M_p = \frac{1}{n}\sum_{1}^{n}(P_t - A_t)^2$$

(iii) Theil's second inequality coefficient:

$$U = \sqrt{\frac{\sum_{1}^{n}(P_t - A_t)^2}{\sum_{1}^{n}A_t^2}}$$

(iv) Inequality Proportions:

(a) Bias proportion: $(\bar{P} - \bar{A})^2 / M_p$

(b) Slope proportion: $(1 - \beta)^2 \sigma_p^2 / M_p$

(c) Disturbance proportion: $(1 - r^2)\sigma_A^2 / M_p$

(d) Variance proportion: $(\sigma_p - \sigma_A)^2 / M_p$

(e) Covariance proportion: $2(1 - r)\sigma_p \cdot \sigma_A / M_p$

σ_p and σ_A are respectively the standard deviations of the predicted and the realized values. β is the regression coefficient in the regression of A_t on P_t, and r is the coefficient of correlation between A_t and P_t.

It should be recalled that since the model-estimates and the extrapolations are estimated with reference to sample periods which are similar to ones for which regression of A_t on P_t is now being undertaken, they are expected to be without 'systematic' errors. This implies that for these two cases $(\overline{P} - \overline{A})$ is expected to be zero and β is expected to be equal to 1. For these cases, therefore, it is only the other parameters which are of interest.

5.1 Tax on Non-Corporate Incomes

The estimated extrapolative model for the revenue from the non-corporate income tax is as follows

(5.1) $$R_t = 1.658 + 1.1095\ R_{t-1}$$

The computed t-value for the regression coefficient in this fit is 17.87, and the coefficient of multiple correlation is 0.98.

Using this equation, benchmark predictions were generated for the period 1961-62 through 1972-73. These predictions are given in Table 5.1 along with the corresponding actual values and the two sets of other predictions, viz., the model-predictions and the budget-estimates.

Values of mean square error, the inequality coefficient and the deviation of the predicted mean from the actual mean were now estimated for the three sets of predictions. These are presented in Table 5.2.

Table 5.1

Revenue from Income Tax on Non-Corporate Assessees: 1961-62 to 1972-73[1]

Year	Actual Receipts	Model-Estimates	Extrapolation	Budget-Estimates
1961-62	165.39	164.63	187.37	133.00
1962-63	185.96	192.78	185.16	163.35
1963-64	258.60	258.73	207.98	218.00
1964-65	266.55	252.48	288.58	250.00
1965-66	271.80	255.85	297.40	294.00
1966-67	308.69	299.61	303.22	292.90
1967-68	325.62	346.93	344.15	290.00
1968-69	378.47	408.45	362.94	319.65
1969-70	448.45	434.58	421.58	362.30
1970-71	473.00	500.89	499.22	436.75
1971-72	537.00	539.08	526.46	491.00
1972-73	602.00	567.52	597.47	583.00

Table 5.2

Prediction Performances: Non-Corporate Income Tax

Forecast	$(\overline{P} - \overline{A})$	Mean Square Errors	Inequality Coefficient
Model-Estimates	0.0	341.5	0.049
Extrapolation	0.0	528.5	0.061
Budget Estimates	-32.30	1680.5	0.109

The value of $(\overline{P} - \overline{A})$ is an estimate of bias in the predictions. It is clear that the budgetary predictions consistently understimate the revenue from income tax on non-corporate assessees. A look at the mean square errors and the inequality coefficients indicates that the estimated models predict better than the simple extrapolation or the budgetary predictions for the given sample period. The performance of the budgetary estimates would seem to be the least satisfactory. However, since there are systematic errors in the budget forecasts whereas such errors are not present

1. Actual figures for 1972-73 are revised estimates.

in the other two sets of predictions, it will be more appropriate to compare the prediction performances after the systematic errors are removed from the budget forecasts. The residual component of mean square error for the budget estimates after the contribution of bias and slope errors is removed is 625.0. Compared to the mean square errors for the model-estimates and the extrapolations, this figure is still higher. Thus, the budget-estimates are still the least satisfactory of the three.

The relative contribution of different sources of errors to the mean square error can be analysed by a study of the inequality proportions.

Table 5.3

Analysis of Mean Square Error: Non-Corporate Income Tax

Forecast	Inequality Proportions				
	Bias	Slope	Disturbance	Variance	Covariance
Model-Estimates	0.0	0.000	1.000	0.005	0.995
Extrapolation	0.0	0.000	1.000	0.008	0.992
Budget Estimates	0.621	0.007	0.372	0.020	0.359

It will be recalled that the variance proportion measures the relative contribution of the error due to incomplete variation. Errors of this nature arise if the forecaster has neglected the causes of fluctuations in the two series. Over time, he is expected to reduce this type of error. On the other hand, covariance proportion measures the relative contribution of the error due to incomplete covariation. Relatively, this type of error stands a smaller chance of correction over time. For the model-forecast and the extrapolations, almost the entire mean square error is due to this latter type of error.

The sizes of the inequality proportions provide an interesting analysis of the budget forecasts. As has already been pointed out, judged from the mean square error, the budget-estimates are inferior to the other two sets of forecasts, even when bias and slope errors are not taken into account. In relative terms, however, bias accounts for the highest proportion of the mean square error. The slope error is relatively small. In the other decomposition, the contribution of the variance error is relatively small. The residual or the disturbance error and the covariance errors are of the type that nothing much can be done about them. But if the bias, i.e., the consistent tendency to underestimate the revenues, can be removed, the budget-estimates can be considerably improved.

If a prediction-realization diagram were to be drawn for the budget estimates, it will be observed that the deviation of the line of regression of A_t on P_t from the line of perfect forecasts is primarily due to the difference in the 'levels' of the two lines rather than due to a difference in the 'slopes'. The former kind of difference arises because the mean point ($\overline{P}_t$, $\overline{A}_t$) does not lie on the L.P.F. The farther away the mean point is from the L.P.F., the bigger is the error due to 'bias'. The regression line for the budget forecast is given by the following.

(5.2) $$A_t = 23.77 + 1.0267 P_t$$

It should be noted that similar regression lines for the model-forecasts and the extrapolations will coincide with the L.P.F. since they do not have a bias or a slope error.

The model- and the budget-forecasts can also be evaluated in terms of relative accuracy analysis. It was suggested in Chapter 3 that the relative mean square error can be used for this purpose. It was defined as

$$RM = M_p / M_x$$

where M_p refers to the mean square error of the predictions in question and M_x refers to the mean square error of the extrapolation.

Mincer and Zarnowitz (1969) suggest the following scale for measuring the quality of a forecast.

$$0 < RM < 1$$

The smaller the value of RM, the better the forecast is. If $RM > 1$, the forecast would be considered to be inferior to the benchmark, although in these cases it would be desirable to consider the ratio of mean square errors for 'linearly' corrected predictions and extrapolations. This is given by

$$RM^c = M_p^c / M_x^c$$

where c refers to 'linearly' corrected forecasts.

For the model predictions, in the present case,

$RM = RM^c = 0.646$

For budget-estimates,

$RM = 3.180$

$RM^c = 1.184$

Accordingly, the budget-forecasts appear to be very unsatisfactory as the relative mean square error, even after corrections for 'systematic' errors, is still greater than one.

Both the model-forecasts and the budget-forecasts may be viewed as being based on two types of information. First, the forecasts utilise the information contained in the past history of the tax-revenue. Secondly, they utilise the information derived from the study of interrelationships of the tax-revenue series with other economic series. It is interesting to analyse the relative contribution of these two types of informations to the generation of a set of forecasts. These relative contributions may be called respectively the 'extrapolative' and the 'autonomous' components of a forecast.

In generating the model-estimates, interrelationships of tax-revenues with other economic variables were explicitly studies and utilised. Hence, the relative contributions of the 'autonomous' component would be expected to be substantial in this case. But for the budget-estimates, it is not clear whether greater significance is attached to the use of one type of information rather than another. This is because the framework and the assumptions behind budgetary predictions are not explicit. In general, the budget estimates for different taxes are provided by the relevant tax-sections of the Ministry of Finance. It will be interesting to analyse indirectly how far the financial experts rely on the past history of the revenue series and how far they incorporate autonomous information.

In Chapter 3, two partial correlation coefficients were suggested for relatively measuring the extent to which (i) the predictive power contained in autonomous information is utilised, and (ii) the predictive power contained in the past history of the predicted variable is <u>not</u> utilised. These measures are the following:

$r^2_{AP\cdot X}$: measure of the net contribution of the autonomous component

$r^2_{AX\cdot P}$: measure of the extent to which predictive power contained in extrapolation is <u>not</u> utilised by the forecast.

Here the subscripts refer to the actual (A), predicted (P) and the extrapolated (X) series. Thus, $r^2_{AP\cdot X}$ is the square of the coefficient of correlation between the actual and the predicted series after the 'linear' effect of the extrapolated series on the actual series has been eliminated.

The values of these squared partial correlation coefficients for the model- and the budget-estimates are given in Table 5.4.

Table 5.4

Utilisation of Autonomous and Extrapolative Information: Non-Corporate Income-Tax

Forecast	$r^2_{AP\cdot X}$	$r^2_{AX\cdot P}$
Model-estimates	0.576	0.344
Budget-estimates	0.167	0.297

A study of $r^2_{AP\cdot X}$ reveals that the model-estimates utilise autonomous information considerably more than the budget-estimates do. In fact, the budget-estimates seem to make little use of the information contained in interrelationships among macro-variables. The non-zero values of $r^2_{AX\cdot P}$ indicate that there was predictive power contained in extrapolative information which was not utilised. Here, the budget-estimates seem to make better use of extrapolative information than the model-estimates.

Finally, the relative weights given to the use of autonomous and extrapolative information were tentatively studied by carrying a multiple regression of budget estimates on model- and extrapolative estimates. It was supposed that the regression coefficient of the model-estimates in this equation would indicate the relative weight given to autonomous information, and that of the extrapolative estimates would give the relative weight attached to the past history of the predicted variable. In fact, the model predictions under-estimate the predictive power of autonomous information and also contain the predictive power of extrapolation. On the other hand, the one-period ahead extrapolation under-estimates the predictive power contained in extrapolative data.

The estimated regression equation of budget-estimates (B) on model-estimates (M) and extrapolation (X) is given by the following:[1]

(5.3) $$B = -16.81 - \underset{(-0.0078)}{0.002047M} + \underset{(3.609)}{.958X}$$

The budget-estimates would seem to primarily rely on the use of extrapolative information. But not much weight can be attached to the estimated coefficients of M as representing the use of autonomous information because it underrepresents it and because it is highly collinear with X.

Overall, the following conclusions may be derived from the above study for the budget- and model-estimates of revenues from the income-tax on non-corporate assessees.

(1) For the given sample period, the budget-estimates perform worse than one-period-ahead extrapolations even when 'bias' and 'slope' errors are removed. The model-estimates perform better.

(2) The budget-estimates consistently underestimate the actual revenues, and the relative contribution of this type of 'bias' error is substantial.

(3) The model-estimates utilise autonomous information much more than the budget-estimates but they relatively underutilize extrapolative information.

(4) The budget-estimates seem to primarily rely on the past history of the revenue-series. The potential information which may

---

1. Figures in brackets are t-ratios.

be derived from a study of the relationships of this revenue-series with other economic series is under-utilised.

5.2 Corporation Tax

For this tax, the following extrapolation model was estimated.

(5.4) $$R_t = 30.99 + 1.02135\ R_{t-1}$$

R refers to the revenue from corporation tax and the subscript refers to the time period. The computed t-value for the regression coefficient in this fit was 8.1 and the coefficient of multiple correlation was 0.93.

The benchmark predictions obtained by this extrapolative model are given in Table 5.5. The corresponding actual values, the model-estimates, and the budget-estimates are also given in this Table.

Table 5.5

Revenue from Corporation Tax: 1961-62 to 1972-73[1]

Year	Actual Receipts	Model-Estimates	Extra-polation	Budget-Estimates
1961-62	156.46	154.66	144.41	141.00
1962-63	221.50	239.72	190.79	178.45
1963-64	274.59	256.66	257.22	227.00
1964-65	314.05	296.81	311.44	306.00
1965-66	304.84	321.26	351.74	371.60
1966-67	328.90	277.27	342.34	372.07
1967-68	310.33	309.70	366.91	350.00
1968-69	299.77	392.54	347.94	320.35
1969-70	353.39	351.58	337.16	326.20
1970-71	370.00	409.85	391.92	342.00
1971-72	472.00	458.71	408.89	411.00
1972-73	558.00	495.09	513.07	493.00

1. Actual figures for 1972-73 are revised estimates.

An analysis of the size and the nature of prediction errors was carried out similar to that for the non-corporate income-tax. Mean square prediction errors and the inequality coefficients for the three sets of estimates can be studied from Table 5.6. The extent of 'bias' for the budget-estimates is also given here.

Table 5.6

Prediction Performances: Corporation Tax

Forecast	$(\bar{P} - \bar{A})$	Mean Square Error	Inequality Coefficient
Model-estimates	0.0	1518.4	0.113
Extrapolation	0.0	1337.0	0.106
Budget-estimates	-10.43	1850.7	0.125

Again, a tendency of underestimation is observed in the case of the budget-forecasts as revealed by the negative value of $(\bar{P} - \bar{A})$. Looking upon the mean square errors and the inequality coefficients, the performance of the budget-estimates still seems to be the least satisfactory. However, for this tax the mean square error for the extrapolative predictions is lower than that for the model-estimates. This implies that the predictive power contained in the past history of the revenue series is not being fully utilised either in the budget-forecasts or in the model forecasts and that there is room for improvement in these methods. But still the model-estimates perform better than the budget-estimates. This is so even when the bias and slope errors are removed from the mean square error of the budget-estimates. The 'linearly corrected' mean square error for this method is then 1736.2. However, for both the model- and the budget-estimates, the value of the relative mean square error computed with reference to the mean square error of the extrapolative benchmark is greater than one. Thus, the performance of these methods is not satisfactory.

The relative contribution of different sources of error in augmenting the mean square is analyzed by an analysis of the inequality proportions.

Table 5.7

Analysis of Mean Square Error: Corporation Tax

Forecast	Inequality Proportions				
	Bias	Slope	Disturbance	Variance	Covariance
Model-estimates	0.0	0.0	1.0	0.041	0.959
Extrapolation	0.0	0.0	1.0	0.035	0.965
Budget Estimates	0.059	0.003	0.938	0.024	0.917

Error due to the incomplete covariation of the predicted series with the realized series seems to be of primary importance in all cases. Error due to incomplete variation seems to be of relatively greater importance for the model-estimates than for the budget-estimates.

For the budgetary predictions, it will again be observed that the relative contribution of bias error is more important than that of the slope error. In a prediction-realization diagram, the line of regression of A_t on P_t will deviate from the line of perfect forecasts not so much because of a different slope but mainly because of a different 'level'. This was also the case with regard to the non-corporate income tax. The estimated line of regression A_t on P_t is given by

$$A_t = 18.57 + .97456\ P_t \tag{5.5}$$

However, the relative contribution of the 'systematic' errors of 'bias' and 'slope' is not substantial compared to that of the disturbance proportion. The implication is that although the budgetary forecasts can be improved by removing the consistent tendency of underestimation, this in itself will not be sufficient. A search is required for causes which affect the revenues from corporation tax and which have not been taken into account in preparing the budget-estimates.

Again, an analysis of how far autonomous and extrapolative information has been utilised in the preparation of different estimates will be useful.

The squared partial correlation coefficients, $r^2_{AP\cdot X}$ and $r^2_{AX\cdot P}$ were calculated for the model- and the budget-estimates. The extent to which a forecast derives from autonomous information is indicated by $r^2_{AP\cdot X}$. On the other hand, $r^2_{AX\cdot P}$ indicates the extent to which predictive power contained in extrapolative information was not utilised. The values for these measures are given in Table 5.8.

Table 5.8

Utilisation of Autonomous and Extrapolative Information: Corporation Tax

Forecast	$r^2_{AP\cdot X}$	$r^2_{AX\cdot P}$
Model-estimates	0.204	0.299
Budget-estimates	0.003	0.230

It may be observed that the budget-estimates rely comparatively less on autonomous information than the model-estimates. In contrast, the model leaves more predictive power contained in the past history of revenue-series unexploited than do the budget-estimates.

As in the case of non-corporate income tax, a multiple regression of budget estimates (B) on model-estimates (M) and extrapolation (X) was carried out to gain some tentative idea of the relative weights attached to autonomous and extraplative informations. It should again be noted that both M and X underestimate the predictive power of autnomous and extrapolative informations. In addition, M also contains the use of extrapolative information. Hence, the interpretation of the coefficients as relative weights is not strictly correct. The estimated equation is given by the following.

(5.6) $$B = 4.526 - 0.264\ M + 1.219\ X$$

The regression coefficients seem to be of a nature similar to what was observed in the case of non-corporate income tax. The importance given to the past history of the revenue series in the preparation of budget-forecasts appears substantial. In contrast, autonomous information is relatively underutilised. This has already been observed from the squared partial correlation coefficients.

In summary, it can be observed that neither the model-estimates nor the budget-estimates have been able to do better than simple one-period ahead extrapolations. An attempt to more fully exploit the predictive power of extrapolative information is warranted.

For the budget-estimates, conclusions similar to those for the non-corporate income tax are derived. First, there is the tendency to underestimate. Secondly, they are primarily dependent upon extrapolative information. However, in the case of corporation tax, the relative importance of 'systematic' error is less compared to that of the 'disturbance' error.

5.3 Import Duties

First, an extrapolative model with a one-period-lag was estimated. This is given by the following equation.

$$(5.7) \qquad R_t = 107.68 + .7509\, R_{t-1}$$

R_t stands for revenue from import duties in period t. The computed t-value of the regression coefficient in this equation is 4.6 and the coefficient of correlation between the dependent and the independent variable is 0.84.

The three sets of predictions, viz., model-, budget-, and benchmark predictions, are given in Table 5.9 along with the corresponding actual revenues from import duties for the period 1960-61 through 1970-71.

Table 5.9

Revenue from Import Duties: 1960-61 to 1970-71

Year	Actual Receipts	Model-Estimates	Extra-polation	Budget-Estimates
1960-61	154.61	169.84	209.67	143.62
1961-62	198.22	189.45	223.77	178.85
1962-63	238.42	274.63	256.52	197.47
1963-64	334.25	390.03	286.71	306.64
1964-65	404.64	339.08	358.66	337.79
1965-66	547.69	463.14	411.52	424.00
1966-67	479.21	507.41	518.93	567.08
1967-68	408.08	404.51	467.51	532.29
1968-69	373.97	466.84	414.10	455.56
1969-70	326.97	292.30	388.49	388.10
1970-71	423.00	361.14	353.20	429.50

For a comparative study of the quality of the different sets of predictions, the deviations between the actual and the predicted means, the mean square errors and the inequality coefficients were again calculated. These statistics are given in Table 5.10.

Table 5.10

Prediction Performances: Import Duties

Forecast	$(\overline{P} - \overline{A})$	Mean Square Error	Inequality Coefficient
Model-estimates[1]	-2.790	2785.4	0.142
Extrapolation	0.00	3845.9	0.167
Budget-estimates	6.531	5117.1	0.193

Again, judged from the size of mean square errors and the inequality coefficients, the budget estimates do not perform well in comparison to the other methods. The model-estimates have the best results. After bias and slope components are removed from the mean square error for the budget estimates, the residual figure is 3596.3. This puts the 'linearly corrected' budget estimates as somewhat superior to the extrapolation. But the model-forecasts are still better for the given sample period.

A ranking of the model- and the budget-estimates is obtained by the relative mean square errors estimated with reference to the extrapolation. For the model-forecast,

$$RM = 0.724.$$

For the budget forecasts, the relative mean square error and the 'linearly corrected' relative mean square error are, respectively, given by the following.

$$RM = 1.33$$

$$RM^c = 0.935.$$

It is worth noting that in contrast to the earlier two taxes, the budget-estimates for import duties have a positive 'bias'. A look at the relevant figures in Table 5.10 indicates that over-estimation has consistently occurred in the post-devaluation period,

1. Since the equation used for predicting these estimates is in a logarithmic form, the estimates are not totally free of bias and slope errors but are nearly so.

and not in the predevaluation period.

Bias, however, does not seem to be an important source of error when the relative contribution of different types of errors are studied with the help of the inequality proportions.

Table 5.11

Analysis of Mean Square Error: Import Duties

Forecast	Inequality Proportions				
	Bias	Slope	Disturbance	Variance	Covariance
Model-estimates	0.003	0.009	0.988	0.022	0.975
Extrapolation	0.0	0.0	1.0	0.088	0.912
Budget-estimates	0.008	0.289	0.703	0.089	0.903

The greater part of the error in all three sets of estimates is observed to be due to incomplete covariation. This type of error has been described by Theil (1966) as relatively more 'hopeless' insofar as the chances of its correction over time are concerned. The first two inequality proportions for the budget-estimates are interesting. In comparison to the non-corporate income tax, here the slope error seems to be more important than the bias error. This implies that the budget-forecasters have not erred so much in anticipating the correct levels of import duties as in anticipating the correct slope of the line of perfect forecasts.

The estimated line of regression of A_t on P_t is given by the following.

(5.8) $$A_t = 95.93 + 0.71544\ P_t$$

A comparison of this with those for the previous two taxes indicates the different nature of error in this case. First, in a prediction-realization diagram, the mean point $(\overline{P}, \overline{A})$ would lie below the L.P.F.

indicating a tendency to overestimate; secondly, the slope of the regression line differs substantially from the line of perfect forecast as compared to the earlier two cases.

For the import-tax revenue also, a study of the relative utilisation of autonomous and extrapolative informations in the budget- and model-predictions may prove useful. The squared partial correlation coefficients are given in Table 5.12.

Table 5.12

Utilisation of Autonomous and Extrapolative Information: Import Duties

Forecast	$r^2_{AP \cdot X}$	$r^2_{AX \cdot P}$
Model-estimates	0.361	0.107
Budget-estimates	0.071	0.007

It is observed from the values of $r^2_{AP \cdot X}$ that the budget-estimates make relatively little use of autonomous information. In contrast, for the model-estimates the autonomous component is relatively more important. On the other hand, judged from the values of $r^2_{AX \cdot P}$, the model-estimates tend to underutilize information contained in the past history of the revenue series relatively more than the budget forecasts.

The regression of budget-estimates on model-estimates and extrapolation is also reported. It should be noted that the model-estimates substantially underrepresent the use of autonomous information. The regression equation is given by the following:[1]

$$(5.9) \qquad B = -131.17 + \underset{(1.09)}{0.1975\, M} + \underset{(5.93)}{1.1936\, X}$$

1. Figures in brackets are the t-ratios

Again, it would seem that the budget-forecasts are primarily dependent on the extrapolative information. Although, for reasons previously given, not much importance can be attached to the coefficient obtained for the model-estimates.

In summary, the following observations can be made regarding the performance of the budget- and model-estimates of revenue from import duties.

(1) For the given sample period, the model-estimates perform better than the budget estimates. This is so even when 'systematic' errors are removed from the budget estimates. This finding is similar to that for the other two taxes. However, in this case, the 'linearly corrected' budget-estimates obtain a relative mean square error which is less than 1. Hence, it would seem that the budget-estimates at least perform better than the extrapolation once systematic errors are removed.

(2) The budget consistently overestimates the revenue from import duties. This is in contrast with the tendency of underestimation reported for the other two taxes. However, overestimation in this case is characteristic only of the post-devaluation period.

(3) For the budget-estimates, 'slope' error contributes more to the mean square error than the 'bias' error. This also is in contrast with the other two taxes, where the latter is relatively more important.

(4) The budget forecasts do not very much utilise autonomous information. The extent to which extrapolative information is not utilised is more for model-estimates than the budget-estimates. This observation is in line with that for the other two taxes.

5.4 Summary and Overall Conclusions

In this study, an attempt was made to provide a theoretical framework for the construction and evaluation of tax-revenue forecasts in a developing economy. In addition, a partial empirical counterpart to this framework was provided in the study of three major Union taxes in India.

The theoretical framework is broader in scope than that suggested by the empirical work. In particular, it provides for building tax-interaction models in economies with multiple tax-rates for various direct and indirect taxes. It was noted in Chapter 2, that revenue estimation in developing economies faces a major problem regarding the way in which discretionary changes in tax-rates and bases should be introduced in the prediction model. Earlier literature in this field suggests two ways of dealing with this problem. One alternative is to construct adjusted revenue series such that the effects of discretionary changes may have been removed from them. We have analysed the limitations of the theoretical assumptions underlying the available adjustment methods. Overall, these methods are not satisfactory. Moreover, predictions based on adjusted revenue data cannot provide information about the effects of tax-policy changes and thus are of limited practical use. The second alternative was to use tax-rates in the regression equations in addition to other exogenous variables. This is theoretically more appealing but has a limited application for taxes with multiple tax-rates for different categories of tax-bases. It may not be possible to use all the relevant tax-rates in the estimation equations because of the implied loss of degrees of freedom for samples of limited size, and because of possible problems of multicollinearity.

In view of these problems it was suggested that tax-rate functions should be estimated so that the rate-structure of a tax can be represented by a limited number of parameters. These parameters can later be used in revenue-estimation equations. Using tax-rate functions, a model specifically allowing for interaction among tax-revenues was developed. It was recognised that the complexities and special characteristics of individual case studies necessitate additional modifications to this framework. Some typical problems in the case of developing economies were identified and suggestions were made regarding possible methods of dealing with them.

It has not been possible to fully incorporate the theoretical considerations in the empirical part of the study. The federal tax-system in India implies that a revenue-estimation model which allows for interaction among tax-revenues can only be achieved when not only Union but major state taxes are taken into account. This would require a study of discretionary changes in these taxes for all the states. In view of the very extensive nature of such a study, this alternative was not pursued.

A comparative analysis of model-estimates and budget-estimates corrected for systematic errors of 'bias' and 'slope' over the given sample periods provide encouraging results inasmuch as the model-estimates consistently perform better. The evaluation of forecasts also provides interesting insights about the quality and nature of budgetary revenue predictions. Overall, the Union budget seems to consistently underestimate tax-revenues and the quality of forecasts can be substantially improved by the removal of systematic errors. Furthermore, the budget-estimates rely primarily on extrapolative information and seem to make little use of autonomous information. This may imply that the prediction work within the

Ministry of Finance tends to be compartmentalized such that the fiscal experts primarily look back upon the past values of the tax-revenues they are concerned with, and do not give appropriate importance to prevalent interrelationships among important macro-variables within the economy. It is also clear that even though the budget predominantly relies upon extrapolative information, it is not able to fully exploit the predictive power contained in it, as even simple one-period ahead extrapolations seem to do better than the budgetary revenue estimates. To confirm these conclusions, however, the prediction-performance of the budgetary estimates should be studied for a bigger sample period, and for a greater number of taxes.

Finally, we may indicate the directions in which the present study could be extended in order to eliminate or reduce the shortcomings of the present work.

First, an attempt could be made towards the construction of a general equilibrium model where the fiscal sector of the economy is interdependent upon other sectors. This would permit determination within the model of important macrovariables previously treated here as exogenous. It would also enable one to obtain greater autonomous information about interrelationships of tax-revenues with other macro- and policy variables.

Secondly, even if a model is specifically developed for the fiscal sector, interaction among tax-revenues could be allowed for.

Thirdly, within the single-equation models, lag-structures in the independent variables, especially that in income, could be explored in greater detail.

Fourthly, an attempt should be made to construct 'composite' forecasts so that the use of predictive power contained in different types of informations may be maximised. It was revealed that the models have not been able to fully utilise the predictive power contained in the past history of tax-revenue series. This suggests work in two directions. First, more powerful extrapolative models than a one period-ahead extrapolation should be developed and secondly model-forecasts should be combined with these forecasts. A weighted average of the two sets of forecasts would be able to more fully exploit the predictive power of autnomous and extrapolative informations. Experiments with constructing 'composite' forecasts in a more general context, such as Bates and Granger (1969) and Nelson (1972), seem to yield encouraging results.

Finally, a comment must be made about actual predictions of revenues for future years. Given estimated prediction-models, forecasts can be obtained simply by feeding in future values of the exogenous variables. In a developing economy, the task of obtaining reliable predictions regarding the exogenous variables proves to be a difficult one. In advanced economies, alternative sources are generally avaialble whereby predictions regarding exogenous variables may be derived. In developing economies such a forecasting 'infra-structure' is not generally available. The quality of forecasts is primarily dependent on the quality of predictions regarding exogenous variables and can never be better than the latter. Because of this interdependence among different forecasts through exogenous variables, the quality of forecasting in developing economies should be considered more as 'evolutionary' in nature rather than isolated and independent.

APPENDIX

List of Tables

Where not otherwise mentioned, the basic data are obtained from the Reports on Currency and Finance, Reserve Bank of India, 1955-56 through 1972-73.

TABLE 1

Data Used in the Estimation of Revenue from Income-tax on Non-Corporate Assessees

Year	Per Capita[1] Net National Product (Rs.)	Year	Population[2] (mid-year) (Millions)	Year	Tax-rate Parameters: First set[3]		Tax-rate Parameters: Second set[4]		Revenue from non-corporate income tax (Rs. Crores)
					r_y^o	r_y^b	r_y^o	r_y^b	
1960-61	306.0	1961	442.21	1961-61	-0.04573	0.05667	0.27141	0.00288	165.39
1961-62	315.2	1962	451.73	1962-63	-0.01773	0.05460	0.29430	0.00267	185.96
1962-63	325.9	1963	461.54	1963-64	0.04818	0.05166	0.34553	0.00250	258.60
1963-64	365.8	1964	471.63	1964-65	0.00129	0.05362	0.30152	0.00272	266.55
1964-65	421.2	1965	482.02	1965-66	0.01353	0.04564	0.27129	0.00228	271.80
1965-66	425.2	1966	492.69	1966-67	0.01514	0.05018	0.29853	0.00251	308.69
1966-67	480.2	1967	503.63	1967-68	0.01514	0.05018	0.29853	0.00251	325.62
1967-68	556.0	1968	514.87	1968-69	0.01496	0.04974	0.29078	0.00256	378.47
1968-69	556.1	1969	526.43	1969-70	0.02742	0.04980	0.30642	0.00252	448.45
1969-70	600.7	1970	538.31	1970-71	-0.01343	0.05937	0.31212	0.00311	473.00
1970-71	633.6	1971	550.24	1971-72	-0.01343	0.05937	0.31212	0.00311	537.00
1971-72	651.1	1972	561.86	1972-73	-0.01343	0.05937	0.31212	0.00311	602.00*

*Revised Estimate

1. Source: Govt. of India (1974): Economic Survey, 1973-74

2. Source: Reserve Bank of India (1972), Report on Currency and Finance, 1971-72

3. From Table 6 of this Appendix

4. From Table 7 of this Appendix

NOTE: In addition to the above series, G.D.P. at current prices was used in the estimation of non-corporate income tax. This series can be read in Col. 1 of Table 2 of this Appendix.

TABLE 2

Data Used in the Estimation of Corporate Tax Revenues

Year	Gross National Product at Current Prices[1]	Year	Super Profits/Sur-Tax Rate	Tax-rate Principal Components[2] (1)	(2)	Corporate tax revenue
	Rs. Crores	Year				Rs. Crores
1960-61	14007.0	1961-62	0.00	-4.40930	-0.53330	156.46
1961-62	14805.0	1962-63	0.00	-2.42363	1.90043	221.50
1962-63	15730.0	1963-64	0.00	-2.42363	1.90043	274.59
1963-64	17977.0	1964-65	0.50	-1.40515	-1.67249	314.05
1964-65	21111.0	1965-66	0.40	-2.18027	-1.85324	304.84
1965-66	21856.0	1966-67	0.40	1.83461	0.03665	328.90
1966-67	25162.0	1967-68	0.35	1.83461	0.03665	310.33
1967-68	29686.0	1968-69	0.35	1.83461	0.03665	299.77
1968-69	30519.0	1969-70	0.25	1.83461	0.03665	353.39
1969-70	33701.0	1970-71	0.25	1.83461	0.03665	370.00
1970-71	36369.0	1971-72	0.25	1.83461	0.03665	472.00
1971-72	38356.0	1972-73	0.25	1.83461	0.03665	558.00*

* Revised estimate

1. Source: Govt. of India (1974), Economic Survey, 1973-74

2. Based on Tax-rates given in Table 5 of this Appendix.

TABLE 3

Data Used in Estimation of Import Duty Revenues and Import Demand Functions

Year	Dummy Variable	Gross National Product at constant 1960-61 Prices	Unit Value of Imports/Wholesale Prices[1]	Volume of Imports[2]	Value of Imports (c.i.f.)	Revenue from Union Excise Duties: Increment Over Last Year	Revenue from Import Duties
		Rs. Crores	1960-61=100	1960-61=100	Rs. Crores	Rs. Crores	
1960-61	0.0	13279.0	100.000	100.00	1139.69	55.70	154.61
1961-62	0.0	13993.0	97.996	95.16	1107.13	72.96	198.22
1962-63	0.0	14796.0	102.120	101.75	1135.57	109.52	238.42
1963-64	0.0	16973.0	98.973	106.19	1222.85	130.75	334.25
1964-65	0.0	19997.0	96.980	114.79	1349.03	71.93	404.64
1965-66	0.0	20624.0	92.185	112.76	1394.05	96.47	547.69
1966-67	1.0	23771.0	62.690	114.32	2078.36	135.86	479.21
1967-68	1.0	28134.0	70.574	124.32	2007.61	114.74	408.08
1968-69	1.0	28808.0	68.080	114.01	1908.63	172.15	373.97
1969-70	1.0	31778.0	68.565	94.30	1567.49	203.64	326.97
1970-71	1.0	34279.0	65.237	93.03	1625.17	234.69	423.00

1. Derived by dividing an index of unit value of imports (1950-51=100) by an index of domestic wholesale prices (1960-61=100) and taking an index with 1960-61 as the base.

2. Derived by dividing value of imports (c.i.f.) by an index of unit value of imports (1950-51=100) and taking an index with 1960-61=100.

TABLE 4

Marginal Income Tax-Rates for Earned Incomes over Rs. 5000 of an Individual in India Read with Appended Notes

No.	Income-Slab (Rs.)	1960-61 & 1961-62	1962-63	1963-64*	1964-65	1965-66	1966-67 & 1967-68	1968-69	1969-70	1970-71, 1971-72 and 1972-73
1	5,000-7,500	.0630	.0735	.12168	.100	.100	.110	.110	.110	.110
2	7,500-10,000	.0945	.1050	.15870	.150	.100	.110	.110	.110	.110
3	10,000-12,500	.1155	.1260	.17844	.150	.150	.165	.165	.187	.187
4	12,500-15,000	.1470	.1575	.20805	.200	.150	.165	.165	.187	.187
5	15,000-17,500	.1890	.2100	.27320	.200	.200	.220	.220	.253	.253
6	17,500-20,000	.1890	.2415	.30218	.200	.200	.220	.220	.253	.253
7	20,000-25,000	.3150	.3465	.39678	.350	.300	.330	.330	.330	.330
8	25,000-30,000	.4200	.3675	.42190	.400	.400	.440	.440	.440	.440
9	30,000-40,000	.4725	.4935	.53908	.550	.500	.550	.550	.550	.550
10	40,000-50,000	.5775	.5985	.63785	.550	.500	.550	.550	.550	.660
11	50,000-60,000	.6300	.6825	.71425	.700	.600	.660	.660	.660	.660
12	60,000-70,000	.6825	.7350	.76150	.700	.600	.660	.660	.660	.770
13	70,000-80,000	.7350	.76125	.78513	.750	.650	.715	.660	.715	.770
14	80,000-100,000	.7875	.76125	.78513	.750	.650	.715	.660	.715	.825
15	100,000-200,000	.8250	.7975	.81275	.825	.6825	.75075	.770	.770	.880
16	200,000-250,000	.8250	.7975	.81275	.825	.715	.7865	.770	.770	.935
17	250,000 and above	.8250	.7975	.81275	.825	.715	.7865	.825	.825	.935

* Not strictly comparable

Notes on Table 4

1. For years 1960-61 through 1964-65 (except 63-64) income-tax was leviable as per a basic tax, a super tax and a surcharge on basic and super taxes. The tax - rate for an income-slab, in these years, is thus given by

(basic rate + super rate) (1 + rate of surcharge)

Only the surcharge relative to earned incomes has been taken into account.

2. In 1963-64 a special additional surcharge was leviable on income as reduced by the income tax paid from it. Hence for this year, the tax-rate is given by

(basic rate + super rate) (1 + rate of surcharge)
+ [1 - (basic rate + super rate) (1 + rate of surcharge)] (rate of additional surcharge)

The income-slabs for the special additional surcharge related to incomes minus income tax paid on it. The writing of the effective tax-rates in the above manner, for the original income-slabs, implies a slight overestimate of tax-rates in the initial categories. For higher income-slabs, the rate of special surcharge did not vary.

3. In 1964-65 the surcharge on earned incomes was leviable only above Rs. 1 lakh.

4. In 1965-66 the basic and super taxes were merged into one. Hence, from now on the effective rates are given by

(income tax-rate) (1 + rate of surcharge)

5. The higher rate of surcharge leviable on earned incomes above Rs. 2.5 lakhs has not been taken into account in 1965-66 and 1966-67. This would have necessitated introduction of another income-slab for this purpose alone.

6. In 1966-67 a flat special surcharge on income-tax (called union surcharge) of 10 per cent was levied. It has continued since then. Hence rates for this and subsequent years are given by

(1 +.1)[(income tax-rate) (1 + rate of surcharge, if any)]

TABLE 5

Tax Rates for Corporate Non-Dividend Incomes in India: Read with Appended Notes

Year	Type of Company / Income							
	D,PS, ≤25,000	D,PS, >25,000, P	D,PS, >25,000, NP	D, PNS,I, first 10 lakhs of income	D,PNS,I, for income in excess of 10 lakhs	D,PNS, NI	F,R	F,NR
	1	2	3	4	5	6	7	8
1961-62	.45	.45	.45	.45	.45	.45	.50	.63
1962-63	.50	.50	.50	.50	.50	.50	.50	.63
1963-64	.50	.50	.50	.50	.50	.50	.50	.63
1964-65*	.425	.45	.50	.55	.54	.60	.50	.65
1965-66*	.425	.45	.50	.45	.54	.55	.50	.65
1966-67	.45	.55	.55	.55	.60	.65	.50	.70
1967-68	.45	.55	.55	.55	.60	.65	.50	.70
1968-69	.45	.55	.55	.55	.60	.65	.50	.70
1969-70	.45	.55	.55	.55	.60	.65	.50	.70
1970-71	.45	.55	.55	.55	.60	.65	.50	.70
1971-72	.45	.55	.55	.55	.60	.65	.50	.70
1972-73	.45	.55	.55	.55	.60	.65	.50	.70

* Not strictly comparable

Key to the classification:

D = Domestic companies
PS = Companies in which public are substantially interested
PNS = Companies in which public are not substantially interested
I = Industrial companies
NI = Companies which are not industrial companies
≤ 25,000 = Companies with income not in excess of Rs. 25,000
> 25,000 = Companies with income in excess of Rs. 25,000
F = Foreign companies
R = That part of a company's income which derives from royalty, fees for technical services and for other specified heads paid by Indian companies.

Thus, for example, D,PS,≤ 25,000 means the category of domestic companies in which public are substantially interested and which earn an income not more than Rs. 25,000 in a year.

Definitions for industrial companies and companies in which public are substantially interested are given in the appended notes.

Notes on Table 5

1. Companies were subject to an income-tax and a supertax in the period 1960-61 through 1964-65. The rate of income-tax was the same for all types of companies and incomes but various exemptions were granted from super tax. The effective rates for these years are derived by adding the rate of income tax to the relevant rate of super tax for the given category.

2. In 1965-66 the income-tax was integrated with super tax and exemptions to various types of companies/incomes were granted from this composite rate. Hence, for 1965-66 onwards the effective rates are the relevant rates for the given categories after account is taken of the exemption provided to each.

3. The rates are presented for a classification of companies/ incomes which was first introduced in 1966-67. For the years 1960-61 through 1963-64, rates can be written for these categories without any difficulty because the classification in these years was that between domestic and foreign companies only as far as taxation of non-dividend incomes was concerned. But a difficulty arises for 1964-65 and 1965-66 because a different and detailed scheme of classification prevailed in these years as compared to the one in 1966-67. As such the rates for categories 4 and 5 in the Table for 1964-65 and 1965-66 are not strictly applicable to the category under which they are written but correspond to a category in close approximation to this under a different system of classification.

4. An industrial company is one which is wholly or mainly engaged in the manufacture or processing of goods or in mining or in the generation or distribution of any form of power. A company is treated as mainly engaged in the above activities if its income from such activities is not less than 51 per cent of total income.

5. Public are substantially interested in a company if (i) it is a company in which at least 40 per cent shares are held (whether singly or taken together) by the Government of India or the Reserve Bank of India or a corporation owned by that Bank, or (ii) if it is not a private company under the Companies Act, 1956 and satisfies a number of prescribed conditions.

TABLE 6

Tax-Rate Functions for Marginal Income Tax-Rates: First Set

An equation of the type

$$r_j = r_y^o + r_y^b\, j$$

was fitted for each year on the effective marginal tax-rates (r_j) given in Table 4, where j varied over income-slabs (j=1, 2, ... 17). r_y^o was interpreted as the tax-rate 'level', and r_y^b as an 'incremental' factor. The least-squares estimates of r_y^o and r_y^b for each year are reported below.

Year	Intercept (r_y^o)	Regression coefficient (r_y^b)	T-Value for the Regression Coefficient	Coefficient of Correlation between r_j and j
1961-62	-0.04573	0.05667	21.71	0.98
1962-63	-0.01773	0.05460	17.52	0.98
1963-64	0.04818	0.05166	17.24	0.98
1964-65	0.00129	0.05362	17.55	0.98
1965-66	0.01353	0.04564	18.20	0.98
1966-67*	0.01514	0.05018	18.17	0.98
1968-69	0.01496	0.04974	18.65	0.98
1969-70	0.02742	0.04980	21.48	0.98
1970-71*	-0.01343	0.05937	24.08	0.99

* The tax-rates for 1967-68 and 1971-72 and 1972-73 were the same as in their respective previous years. The critical value of t for a 95% confidence interval for 15 degrees of freedom is 1.753. All the regression coefficients are significantly greater than zero. The estimates of r_y^o and r_y^b are then used in the estimation of revenue equation for non-corporate income-tax revenue.

TABLE 7

Tax-Rate Functions for Marginal Income Tax Rates: Second Set

An equation of the type

$$r_j = r_y^o + r_y^b m_j$$

was fitted for each year on the effective marginal tax-rates (r_j) given in Table 4 of this Appendix, where m_j refers to the mid-point of each income-slab. r_y^o and r_y^b are interpreted as tax-rates 'levels' and 'incremental factors', respectively. Their estimates are given below.

Year	Intercept (r_y^o)	Regression coefficient (r_y^b)	t-Value for the Regression Coefficient	Coefficient of Correlation between r_j and m_j
1961-62	0.27191	0.00288	4.79	0.78
1962-63	0.29430	0.00267	4.32	0.74
1963-64	0.34553	0.00250	4.19	0.73
1964-65	0.30152	0.00272	4.68	0.77
1965-66	0.27129	0.00228	4.54	0.76
1966-67*	0.29853	0.00251	4.54	0.76
1968-69	0.29078	0.00256	4.91	0.79
1969-70	0.30642	0.00252	4.76	0.78
1970-71*	0.31212	0.00311	5.27	0.80

* The rates for 1967-68 were the same as in 1966-67. The rates in 1971-72 and 1972-73 were the same as in 1970-71.

The coefficients are seen to be significantly greater than zero when the estimated t-values are compared with the critical value of t = 1.753 for a 5% significance level for 15 degrees of freedom.

TABLE 8

Matrix of Coefficients of Correlation between Tax-Rates for Different Categories of Corporate Assessees and/or Incomes[1] and their Principal Components

	Categories of Tax-Rates[1]							Principal Components	
	1	2	3	4	5	6	8	P.C.₁	P.C.₂
1	1.000	0.146	-0.144	0.020	-0.367	-0.440	-0.382	-0.251	0.965
2		1.000	0.911	0.947	0.819	0.771	0.859	0.903	0.395
3			1.000	0.935	0.973	0.941	0.931	0.984	0.095
4				1.000	0.872	0.870	0.876	0.942	0.271
5					1.000	0.986	0.960	0.981	-0.135
6						1.000	0.953	0.970	-0.210
8							1.000	0.976	-0.131
P.C.₁								1.000	0.000
P.C.₂									1.000

1. These categories are defined in Table 5 of this Appendix. Category 7 is not included in this table or in the computation of the principal components as its variance is zero.

BIBLIOGRAPHY AND REFERENCES

Auld, D.A.L. (1970), 'Fiscal Marksmanship in Canada', Canadian Journal of Economics, 3(3), 507-11.

Bahl, R.W. (1972), 'A Representative Tax System Approach to Measuring Tax Effort in Developing Countries', Int. Mon. Fund Staff Papers, 24(1), 69-79.

Bates, J.M. and Granger, C.W.J. (1969), 'The Combination of Forecasts,' Oper. Research Quarterly, 20, 451-68.

Berney, R.E. (1970), Tax Structure Variations in the State of Washington, Pullman, Washington.

Berney, R.E. and Frerichs, B.H. (1973), 'Income Elasticities for State Tax Revenues: Techniques of Estimation and Their Usefulness for Forecasting', Public Finance Quarterly, 1(4), 409-425.

Bernier, C.D.P. (1968), 'Forecasting the Yield of Federal Taxes', Canadian Tax Journal, 16, 491-512.

Blackburn, J.O. (1967), 'Implicit Tax Rate Reductions with Growth, Progressive Taxes, Constant Progressivity, and a Fixed Public Share', American Economic Review, 57(1), 162-68.

Borkar, V.V. (1971), Income Tax Reform in India, Popular Prakashan, Bombay.

Box, G.E.P. (1966), 'Use and Abuse of Regression', Technometrics, 8(4), 625-29.

Chelliah, R.J. (1971), 'Trends in Taxation in Developing Countries', Int. Mon. Fund Staff Papers, 18.

Cutt, J. (1969), Taxation and Economic Development in India, Praeger, New York.

Davies, D.G. (1960), 'Progressiveness of a Sales Tax in Relation to Various Income Bases', Amer. Econ. Review, 50, 987-95.

Due, J.F. (1968), Government Finance: Economics of the Public Sector, Richard D. Irwin Inc., Homewood.

Due. J.F. (1970), Indirect Taxation in Developing Economies: The Role and Structure of Customs Duties, Excises, and Sales Taxes, John Hopkins, London and Baltimore.

E.C.A.F.E. (1968), Feasible Growth and Trade Gap Projections in the E.C.A.F.E. Region, Bangkok.

Gandhi, V.P. (1970), Some Aspects of India's Tax Structure: An Economic Analysis, Vora, Bombay.

Ghai, D.P. (1966), Taxation for Economic Development: A Case Study of Uganda, East African Studies 23, East African Publishing House, Kampala.

Goldberger, A.S. (1964), Econometric Theory, Wiley, Chichester.

Goode, R. (1964), The Individual Income Tax, The Brookings Institution, Washington, D.C.

Govt. of India (1974), Economic Survey, 1973-74, New Delhi.

Granger, C.W.J. and Newbold, P. (1973), 'Some Comments on the Evaluation of Economic Forecasts', Applied Economics, 5, 35-47.

Grove, E.W. (1969), 'Econometricians and the Data Gap: Comment', Amer. Jour. of Agr. Econ. 51(1), 184-87.

Groves, H.M. and Kahn, C.H. (1952), 'The Stability of State and Local Tax Yields', American Economic Review, 42, 87-102.

Hansen, R. (1962), 'An Empirical Analysis of the Retail Sales Tax with Policy Recommendations', National Tax Journal, 15, 1-13.

Harris, R. (1966), Income and Sales Taxes: The 1970 Outlook for States and Localities, The Council of State Governments, Chicago.

Hey, J.D. (1974), Statistics in Economics, Martin Robertson, London.

Hunt, S.J. (1971), 'Distribution, Growth, and Government Economic Behaviour in Peru' in Gustav Ra is (ed.), Government and Economic Developments, Yale University Press, New Haven.

Johnston, J. (1972), Econometric Methods, McGraw-Hill, New York.

Klein, L.R. (1968), An Essay on the Theory of Economic Prediction, Yrjo Jahnssonin Saatio, Helsinki.

Legler, J.B. and Shapiro, P. (1968), 'The Responsiveness of State Tax Revenues to Economic Growth', National Tax Journal, 21, 46-56.

Levin, J. (1968), 'The Effects of Economic Development on the Base of a Sales Tax: A Case Study of Colombia', Int. Mon. Fund Staff Papers, 15, 30-101.

Lotz, J.R. (1970), 'Measuring the Effects of Fiscal Policy in Developing Countries', Finanzarchiv, 29(2), 238-55.

Mansfield, C.Y. (1972), 'Elasticity and Buoyancy of a Tax System: A Method Applied to Paraguay', Int. Mon. Fund Staff Papers, 19(2), 425-46.

Mincer, J. and Zarnowitz, V. (1969), 'The Evaluation of Forecasts', in Mincer, J. (ed.), Economic Forecasts and Expectations: Analysis of Forecasting Behaviour and Performance, NBER, New York.

Mishan, E.J. and Dicks-Mireaux, L.A. (1958), 'Progressive Taxation in an Inflationary Economy', Amer. Econ. Review, 48, 590-606.

Morawetz, D. (1971), 'The Sensitivity of the Yield of Personal Income Tax in the United Kingdom: A Note', Economic Journal, 81, 612-616.

Musgrave, R.A. (1969), Fiscal Systems, New Haven, London.

Mushkin, S.J. and Lupo, G.C. (1967), 'Project '70: Projecting the State and Local Sector', Review of Econ. and Stat., 42, 237-245.

Nelson, C.R. (1972), 'The Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy', American Economic Review, 62, 902-917.

Pechman, J.A. (1965), 'A New Tax Model for Revenue Estimating' in Peacock, A.T. and Hauser, G. (eds.), Government Finance and Economic Development, O.E.C.D., Paris.

Prest, A.R. (1962), 'The Sensitivity of the Yield of Personal Income Tax in the United Kingdom', Economic Journal, 72, 576-96.

Ray, C.L. (1966), 'A Note on State Tax Stability Criteria', National Tax Journal, 19, 207-209.

Reserve Bank of India (Yearly), Report on Currency and Finance, 1955-56 through 1972-73, Bombay.

Sahota, G.S. (1961), Indian Tax Structure and Economic Development, Asia Publishing House, London.

Schaefer, J.M. (1969), 'Sales Tax Regressivity Under Alternative Tax Bases and Income Concepts', National Tax Journal, 22(4), 516-527.

Schwartz, L. (1970), 'An Econometric Model of Government Revenues: Import Duties', Applied Economics, 2(2), 133-39.

Shaw, G.K. (1972), Fiscal Policy, MacMillan Studies in Economics, London: Basingstoke.

Shourie, A. (1972), 'The Use of Macro-economic Regression Models of Developing Countries for Forecasts and Policy Prescription: Some Reflections on Current Practice', Oxford Economic Papers, 24(1), 1-35.

Singer, N.M. (1968), 'The Use of Dummy Variables in Estimating the Income-Elasticity of State Income-Tax Revenues', National Tax Journal, 21, 200-204.

Singer, N.M. (1970), 'Estimating State Income Tax Revenues: A New Approach', Review of Econ. and Stat., 52, 427-433.

Soltow, L.R. (1955), 'The Historical Rise in the Number of Tax-payers in a State with Constant Tax Law', National Tax Journal, 8, 379.

Streissler, E.W. (1970), Pitfalls in Econometric Forecasting, Institute for Economic Affairs, London.

Theil, H. (1971), Economic Forecasts and Policy, North Holland, Amsterdam (2nd Ed.).

Theil, H. (1966), Applied Economic Forecasting, North Holland, Amsterdam.

Theil, H. (1967), Economics and Information Theory, North Holland, Amsterdam.

Theil, H. (1971), Principles of Econometrics, North Holland, Amsterdam.

U.N.C.T.A.D. (1968), Trade Prospects and Capital Needs of Developing Countries, U.N., New York.

Waldorf, W.N. (1967), 'The Responsiveness of Federal Income Taxes to Income Change', Survey of Current Business.

Walters, A.A. (1970), An Introduction to Econometrics, McMillan, London.

Wilford, W.T. (1965), 'State Tax Stability Criteria and the Revenue-Income Elasticity Coefficient Reconsidered', National Tax Journal, 18, 304-312.