CHAPTER - IV
DATA BASE AND METHODOLOGY

The importance of any of the empirical studies is generally examined and valued by its database and methodology. The analysis on database, in general, includes - (i) the choice and selection of the period of the study (ii) definition and recognition of the variables on which the data are collected, (iii) methods and sources of data collection, (iv) problems and limitations of the collected data and finally (v) adjustments, if necessary, for the data. All these issues on database have been covered in the first four sections of this chapter. This chapter also includes the methodology of the study. Methodology of a study comprises proper choice and selection of the methods for the analysis and interpretation of the data set, consistent with the objectives of the study. The last three sections of this chapter discuss the methods that are used in this study for proper evaluation of the performance of the life insurance in India during pre and post reforms period and also for identifying the probable causes behind the observed performance of LICI.

4.1 Period of the Study

For the analysis of growth and stability of different indicators of life insurance business of Life Insurance Corporation of India and also to measure its performance in terms of different ratios, the period chosen for the study is from the financial year 1962-63 to 1999-2000 (i.e., total 38 years). This whole period has been initially divided into two sub-periods (normally, from 1962-63 to 1979-80 and from 1980-81 to 1999-2000) and again divided into four sub-periods representing separately four decades (viz, 1960s, 1970s, 1980s, and 1990s) to facilitate the comparative analysis of the performance of LICI in different sub-periods as well as different decades of our study period. But in case of identifying the impact of different socio-economic factors on the Indian life insurance business from time series analysis and also for others (which
have been spelt out in the relevant portions of the text) we are forced to consider relatively small sample periods depending on the availability of data. Likewise, due to non-availability of uninterrupted time series data at the state level, we have made cross sectional study at three points of time, *viz.*, 1970-71, 1980-81 and 1990-91 to determine the effects of socio-economic factors on the performance of life insurance business of Life Insurance Corporation of India.

Regarding the selection of time period one question may arise : why has 1962-63 been chosen as the starting year inspite of the availability of data since nationalisation (i.e. 1957)? 1962-63 has been chosen as the starting year on the following grounds. The Corporation had decided to change its accounting year from the calendar year to the financial year from 1962-63 and this method of selecting accounting year is still continuing. So to maintain the uniformity of periodicity of accounting year 1962-63 is taken as the starting year. Further to assess the performance of LICI after its nationalisation some initial years should be left out because the impact of nationalisation might not be started instantaneously from the year 1956-57. On the same ground one may raise objection to the choice of 1980-81 as the cut-off point for the sub-periods because the effect of recommendations made by Era Sezhiyan Committee regarding the review of the working of LICI might not be readily reflected from the very financial year, 1980-81 on which it submitted its report. It may be considered as apparently true. But from the financial statements placed by LICI in its annual reports from 1980 – 81 onwards and from the personal interview of the life insurance personnel, it is depicted that from the very initial period of reforms suggested by Era Sezhiyan Committee, LICI did better performance.

4.2 Data Base for Measuring Growth and Stability

The whole analysis of the study is based on the secondary data which have been collected from various recognised official publications. The basic data that have been collected for the measurement of growth and stability in the performance of life insurance business of Life Insurance Corporation of
India for the period 1962-63 to 1999-2000 are: number of policies of new individual business (in lakhs), total sum assured in new individual business (Rs. in crores), premium income from new individual business (Rs. in crores), total sum assured in in-force individual business (Rs. in crores), number of policies of in-force individual business (in lakhs), premium income from in-force individual business (Rs. in crores), number of policies of new individual business within the country (in lakhs), total sum assured in new individual business inside the country (Rs. in crores), premium income from new individual business (Rs. in crores), number of policies of individual in-force business within the country (in lakhs), total sum assured in in-force individual business inside India (Rs. in crores), premium income from inside in-force individual business (Rs. in crores), number of policies of new individual business outside India (in lakhs), total sum assured in new individual business outside the country (Rs. in crores), premium income from new individual business (Rs. in crores), number of policies of in-force individual business outside India (in lakhs), total sum assured in in-force individual business outside India (Rs. in crores), and premium income from in-force individual business outside India (Rs. in lakhs). During the same period we have also collected data for investment (total investment, sector-wise investments and zone-wise stock exchange investments), distribution of individual business between urban and rural areas [both in terms of number of policies (in lakhs) and total sum assured (Rs. in crores)], individual items of income and outgo of LIC, particulars of insurance personnel and number of zone-wise divisional and branch offices of LIC. Data regarding zone-wise breakup of new and in-force individual business [in terms of number of policies (in lakhs), total sum assured (Rs. in crores), and premium income (Rs. in crores)] and group life insurance (including group gratuity) and superannuation business (both new and in-force) have been collected for the period 1970-71 to 1999-2000. As per availability, the data regarding product-
wise insurance business and pension and annuity business have been collected for the period 1980-81 to 1999-2000.

The sources of these data are (i) different Annual Reports of the Life Insurance Corporation of India, (ii) Report of the Committee to Review the working of LIC (September 1980), (iii) different issues of LIC's diary and (iv) Life Insurance Compendium 1999-2000. In addition, we have used the wholesale price index published through the 'Hand Book of Statistics on Indian Economy' by the Reserve Bank of India (2001) to transform all the nominal variables in their real terms.

Information regarding total individual business (both inside and outside the country) for the financial years 1962-63 to 1965-66 are inclusive of the business secured under group and superannuation scheme, as separate figures for group business are not available. It should be worth mentioning that up to 1965-66, sum assured in group business and also premium income from group business were very insignificant in relation to the total business of LIC. Further, the general practice of LIC during that period was that it had no intention to expand its business outside the territory of India very much.

Data for the distribution of new inside individual business between rural and urban areas, however, are collected up to the financial year 1979-80 from the Report of the Committee to Review the working of LIC and for the rest period collected from the Annual Reports of LIC.

Due to the uncertainty and unreliability of securing detailed information regarding the data on group business prior to the year 1970-71 and due to the very negligible volume of group life insurance business in respect to total life insurance business of LIC up to the year 1970-71, the group and its other related information are brought under discussion from 1970-71 onwards. For the period 1970-71 to 1976-77, the data have been collected from the Report of the Working Committee to Review the working of LIC and thereafter these are collected from the Annual Reports of LIC. It should be mentioned in this connection that from 1988-89 onwards the
annuity of in-force business included also the annuity obtained from the group superannuation (cash accumulation) schemes. In the years 1988 - 89 and 1989-90, number of lives and total sum assured included those of IRDPGI beneficiaries. In all these cases necessary adjustments have been made to maintain the data comparable and consistent over time as far as possible.

There were only four zones (namely, Eastern, Western, Northern, and Southern) at the very beginning of the period (i.e., 1962-63) of our study. Later on, these four zones were reorganised into seven zones by forming additionally Central, North Central and Southern Central zones. But to maintain parity in our study we have made the following adjustment keeping initial four zones intact. From 1979-80 onwards in case of Western region a state of the Central zone, namely M.P. was added to it and another state of the Central zone, namely, U.P. was added to the Northern Zone. Again from 1990-91 onwards, Southern Central zone has been included to the Southern region.

In general, data on total premium income includes first year premium as well as renewal premium in case of both individual and group business. To maintain consistency with the earlier figures, from 1981-82 onwards, first year premium income and renewal premium income obtained from the published revenue account of LICI, have been added to get the amount of total premium income. Further, from the year 1996-97 the premium income of 'Jeeban Suraksha' business was included along with the total premium income in the annual reports. But as the nature of 'Jeeban Suraksha' policy is different, so its premium income has been excluded from total premium income since 1996-97 to keep the data comparable over time. To get the data of the premium income of new individual business during the period 1962-63 to 1979-80, premium income for group insurance and superannuation annual premium receivable have been deducted from the first year premium income as those were published in combined form. However, separately individual premium income and premium for group life insurance business were obtained directly

Gross income of LICI in our study includes total premium income (renewal as well as first year premium income), single premium and other income from the consideration of annuity, investment income and other receipts. Income from investment means gross income without deducting income tax. As mentioned earlier, since 1996-97 income from 'Jeeban Suraksha' business has also not been included in computing total income of LICI.

Data regarding total agent and active agent and other life insurance personnel have been collected from the Report of the Committee to Review the Working of LICI (1980) up to 1979-80 and thereafter from the respective annual reports of LICI. Here active agent means an agent who has done some business during the current year irrespective of the number of policies or amount of sum assured.

4.3 Data Base for the Ratio Analysis and for Identifying the Socio-Economic Determinants

For computing the relevant ratios (to be mentioned in section 4.6) applied for performance appraisal of life insurance business in India during the study period and also to measure the impact of different socio-economic factors on the Indian life insurance business, apart from the basic data related to the life insurance business (as mentioned in the earlier section), the other necessary data which have been collected are gross domestic product, gross domestic savings, household savings, public sector employment, population, total life fund of LICI, organised private sector employment, total investment of LICI, total surplus of LICI, total premium income, total management expenses, expenses for new business and renewal expenses for the country as a whole and also for her 13 major states [viz., Andhra Pradesh, Assam, Bihar, Gujarat, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Orissa, Rajasthan,
Tamil Nadu, Uttar Pradesh and West Bengal) if of course, relevant state-wise data are available. Regarding the demographic factor, we have collected total population, proportion of male population, proportion of urban population, literacy ratio, proportion of able-bodied persons, average size of household, etc. Here household means a group of persons who normally live together and take their meals from a common kitchen unless the exigencies of work prevent any of them from doing so and able-bodied persons means the population group belonging to the age group 15 years to 59 years (i.e., who are able to earn income for his family). For the country as a whole all these data have been collected over time during the study period. But for the 13 major states of India, which have been selected for the classified analysis of the study, few of these data are time series data and most of others have been collected on three time points, namely 1970-71, 1980-81 and 1990-91 (all have been spelt out in Chapter VII).

The sources for these data are (i) different publications of the annual report of LICI, (ii) Report of the Committee to Review the working of LICI (September 1980), (iii) LICI Compendium, 1999-2000, (iv) different issues of LIC Diary, (v) Hand Book of Statistics of Indian Economy, (vi) Statistical Outline of India: 2000-2001, Tata Services Limited, (vii) Economic Survey of India (various issues), (viii) Statistical Abstract of India, (various issues), (ix) Currency and Finance, RBI Bulletin (various issues), (x) Statistical Abstract of West Bengal (various issues), (xi) Census Atlas, (xii) Primary Census Abstract, General Population (various issues), (xiii) Census of India and State Profile (various issues), (xiv) Primary Census Abstract, Social and Cultural Tables (various issues) and (xv) Primary Census Abstract, General Economic Tables (various issues). Various sources have been consulted to check the accuracy and consistency of data.
4.4 Limitations of the Data

There are few common and unavoidable (though sometimes negligible) general problems in collecting secondary data which are required for our study. These problems are related to the (i) error of rounding off decimal figures, (ii) printing errors, (iii) publication of data with different objectives, (iv) sometimes publication interpolated or provisional rather than actual figures, etc. Apart from the general problems, some specific problems have also arisen with the method of data collection in the study. At first, data relating to premium income has not properly been segregated as first year’s premium, first year’s premium from individual business, first year’s premium from group life insurance business, individual renewal premium and renewal premium from group business in the annual reports of LIC through out the study period, though those are very much required for the classified analysis. We have tried to segregate roughly the premium income of group business from the premium income of individual business. For instance, from 1965-66 onwards, first year’s premium income from group business has been calculated by subtracting annual premium of total new individual business, from the total first year’s premium income. In this sort of calculation, lapses or revivals of the existing policies have not been properly taken care of. It should be mentioned in this connection that before 1965-66 group insurance premium income was very much insignificant in respect of total premium income, which has been taken as individual premium income in the study.

To assess the zone-wise performance of LIC, we have noticed that in 1962-63 there were four zones, namely, Eastern, Western, Northern and Southern zone. But over time, specially after 1979-80, the number of zones changed into five and the annual report of 1979-80 showed another zone, known as Central Zone. Again in 1990-91, Southern zone was divided into Southern Central Zone and Southern zone as a result of which total number of zones increased to six. In 1992-93 again Central zone was divided into North Central zone and Central zone. So at present there are seven zones of LIC.
But to maintain uniformity in the analysis for the whole study period, as mentioned earlier, we have taken into consideration the number of zones as four (namely, Eastern, Western, Northern and Southern). So from 1962-63 the states which have been separated to create additional zones have been again taken into the original zones and accordingly all the related data have been adjusted. For example, the Central Zone’s one state, M.P. is added to the Western zone and the Central zone’s another state, U.P is added to the Northern zone. Further, the Southern Central zone is wholly added to the Southern zone. In such type of reorganisation and data adjustment perfect accuracy could not be restored because in some zones actually whole part of a particular state has not been included, rather a state is divided to include with more than one zones.

It has been quite impossible to collect data regarding the distribution of policy-holders according to activity status from the annual reports of LIC. The total value of policies (to be measured either in terms of sum assured or in terms of premium income) taken by a person at a certain level (or range) of income is also not available. Further, the total number of persons under the insurance net are also not reported by LIC, though information is given on total number of policies. But total number of policies and total number of persons taken life insurance policies are not same (many persons have more than one policy). Besides these, the data on urban-rural division of in-force individual life insurance business could not be collected from any of the data sources. But all these data would be required if one wishes to make a detailed study.

To know the impact of socio-economic factors on the life insurance business, state-wise time series analysis has been done by taking only two explanatory variables due to non-availability of data on more variables. In order to include more independent variables, however, regression equations have been fitted cross sectionally (over the selected major states in India) only on three time points (viz., 1970-71, 1980-81 and 1990-91) separately. We could
not be able to include other necessary time points namely, 1962-63 and 1999-2000 in the regression analysis due to the lack of reliable data.

For each of the major states, total sum assured in individual life insurance business is determined by adding the sum assured of different divisions situated in the territory of that state. But the area of a particular division may be the territory of more than one state. Therefore, it has been difficult to get actual state-wise figures on sum assured in individual life insurance business.

Inspite of all these limitations in the data base (which of course, could not be totally avoided in any of the empirical studies), the best effort has been given here to transform the data base as far reliable as possible and to make the analysis qualitatively classified as per the objectives of the study.

4.5 Methodology for the Measurement of Growth and Stability of Different Performance Indicators of LICI

For the measurement of growth and stability of different performance indicators, the researchers have used different methods in their studies, time to time, improving the earlier existing methods. The improvements of the methods are often observed in the fields, like, agriculture, industry or any specific part of the sector of an economy. We shall use the latest and sophisticated methods of measuring growth and stability [as prescribed by Boyce (1987) and Chattopadhyay (2001)] for evaluating the performance of life insurance sector in India. Along with the estimation of growth and stability of the performance indicators, their levels are also calculated in terms of arithmetic mean, because a high stable growth is required for a variable which is relatively at a higher level (that would be represented through its value of arithmetic mean). Before calculating growth or stability or arithmetic mean, the performance indicators which are expressed in nominal terms have been first transformed in their real terms. For this transformation of nominal variable into real one, the corresponding series of the nominal variable has
been deflated by the series of wholesale price index. In this way, the effect of inflation will also be avoided from the measurement of the performance in terms of growth, stability and mean. The performance of LICI will be evaluated by measuring growth, stability and mean of different components related to its business, investment, income-outgo, insurance personnel, offices, insurance products, etc. In the following three sub-sections, the actual choice of the methods for measuring whole period's growth rate, sub-period's growth rates and also for measuring stability have been explained.

4.5.1 Methods of Measuring Whole Period's Growth Rate

The subject of evaluating performance through measurement of growth rates of the selected performance indicators over time is a very tricky area, because one can statistically manipulate the estimated growth rates by the arbitrary choice of the period of the study (specially by suitably selecting the few initial and end years of the study), faulty selection of the statistical techniques, avoidance of econometric problems in estimation, etc. In our study all these issues have adequately been taken care to obtain the accurate estimates of the growth rates.

As mentioned in the first section of the chapter, the whole period of the study is from 1962-63 to 1999-2000, by which the few initial years after nationalisation of LICI (during this initial years most probably the parameters for measuring performance were at low levels) have been excluded and both the periods before and during the economic reforms have been included. With these considerations the estimation of the growth rates would be more or less free from errors. Very low or high value of a variable during few initial or end years may lead to erroneous results in growth estimation. But in our study, through the appropriate selection of the whole period, estimation of growth rates would be more or less free from error and also our study will be able to accommodate all the periodic performance changes (including both pre and post reforms periods).
In most of the research studies in the areas of business and commerce, growth rate from a time series is mainly calculated either by comparing two arbitrarily chosen points or by comparing the averages of same arbitrarily chosen periods at the beginning and end of the series or by taking the average of all point to point growth rates \[(y_t - y_{t-1})/y_{t-1} \text{ for all values of } t = 2, 3, \ldots n\]. But these methods were highly criticised by Vidya Sagar (1980). The accurate method of obtaining measure of growth rate is to postulate a hypothetical trend function and to estimate statistically its parameters (by applying ordinary least square method) which offer the actual measure of growth rate over the period of a series. In this normal statistical procedure, however, one problem crops up. This is the problem of choice of appropriate trend equation.

Empirically, it has been observed that the popular forms of trend equation for growth estimation are linear \((y_t = a + bt)\), parabolic \((y_t = a + bt + ct^2)\), exponential \((\ln y_t = a + bt)\), log quadratic \((\ln y_t = a + bt + ct^2)\), higher degree polynomial \((y_t \text{ or } \ln y_t = a_0 + b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + \ldots)\) and S-type growth curves (Gompertz, Logistic, etc.). But each of these trend equations has its own interpretation and restriction upon the character of the growth process. For example, linear trend implies constant absolute growth [which can be divided by the harmonic mean of the series, as argued by Minhas (1966), to get average compound growth rate]. Similarly, exponential or log-linear trend \([\ln y_t = a + bt]\) implies constant compound growth rate, \(b\) [however, the correct formulation is \(e^b - 1\), as pointed out by Dandekar (1980)]. Although parabolic and log – quadratic trends are more flexible than linear and exponential as they permit non-zero second derivative with respect to time (implying acceleration or deceleration in the growth process), they, nevertheless, impose restriction that the second order derivatives are constant [In this connection it is to be noted that in the case of second degree polynomial equation as \(t\) and \(t^2\) are correlated, time has to be normalised (shifting origin to the mid-point of the series) for the avoidance of multicollinearity problem, as argued by Reddy.]
One way to allow a non-constant second derivative is to add a cubic term of time (i.e., $t^3$) in the trend equation [as used by Ray (1983)]: but this process is unending. Further, by doing so, the very objective of growth measurement would either be violated (having more than one turning points in the growth curve of polynomial of degree three or more) or be meaninglessly laborious. Like log-quadratic trend, S-type growth curves [applied by Rudra (1970), Dey (1975), Reddy (1978), Das (1978), Chattopadhyay (2001), etc.] are also open to the criticism regarding the behaviours of the acceleration/deceleration in the growth process. Further, the results of non-linear estimation procedure [either by ‘Group Average Method’ or by transformation approach (as prescribed by Reddy and Chattopadhyay) or by search technique] in case of S-type growth curves are non-comparable with the results of the OLS method, applied in the cases of other trend equations, as some of these procedures of non-linear trend estimation are arbitrary and defective and for others the relative merits of the estimation method (vis-à-vis others) have not been established by the Monte-Carlo experiment. So in growth measurement exercise the choice of trend equation from amongst the available alternatives is very crucial.

Inspite of these limitations, as trend fitting exercise is the appropriate method of measuring growth rates (of performance indicators) one has to select a more or less suitable trend equation among the available alternatives. As the span of the whole period of our study is relatively large (38 years), so we have to select such a trend equation which help us to estimate directly the growth rate as well as the acceleration/deceleration, if any, in the measured growth rate. To accommodate all these features we have selected log-quadratic trend equation to estimate whole period’s growth rate. More specifically, by estimating the trend-equation, $\ln y_t = a + bt + ct^2$ over the whole period of a series and by normalising the time (i.e., shifting origin to the midpoint of the whole period, i.e., by taking $t = 0$ for the mid of the year 1980), the estimated value of $b$ gives us the measured growth rate of the performance.
indicator for the whole period and the estimated value of \( c \) denotes whether there is acceleration (when \( c \) is significantly positive) or deceleration (when \( c \) is significantly negative) or constancy (when \( c \) is insignificant) in the estimated growth rate. Further, this sort of estimation of the parameters (\( b \) and \( c \)) is free from multicollinearity problem as the explanatory variables \( t \) and \( t^2 \) of log quadratic trend equation remain uncorrelated due to normalisation of time.

To test the 'goodness of fit' of the estimated trend equation we shall use the coefficient of determination adjusted by the degrees of freedom which is popularly symbolized as adjusted \( R^2 \). A high value of adjusted \( R^2 \) or alternatively statistically significant \( F \) value (as adjusted \( R^2 \) is statistically tested by \( F \) distribution) implies that the trend equation is fitted well for the given time series data. On the other hand, a very low value of adjusted \( R^2 \) or statistically insignificant \( F \) value denotes the poor fitting of the trend equation and this will happen when there is no growth in the series of the concerned variable.

In the estimation of trend growth rate from a time series data, a most probable econometric problem in the disturbance term arises and this problem is known as auto-correlation problem. In the earlier research studies in this area for the detection of auto-correlation problem the estimated value of DW (Durbin – Watson) statistic has been reported. But reporting the estimated value of DW statistic in the tables has been mostly done very customarily in the sense that there are very few research studies in which auto-correlation problem in the disturbance term has been corrected when the estimated value of DW statistic establishes such necessity. But in our study we have not only reported the estimated value of the DW statistic along with the estimated growth, but also corrected auto-correlation problem in case of its presence initially established by the DW statistic in the disturbance term. More specifically, the auto-correlation problem in our study has been corrected by

'Cochrane–Orcutt two step procedure' signifies that initially on the original time series data the chosen trend equation (in our case log - quadratic trend) is to be estimated. Then the predicted ($\hat{y}_t$) and the residual ($e_t$) values are to be computed. In case of the presence of auto-correlation problem, detected by the DW statistic, either the first order ($e_t = \rho e_{t-1} + \varepsilon_t$) or second order ($e_t = \rho e_{t-1} + \gamma e_{t-2} + \varepsilon_t$) auto-regressive scheme is to be estimated on the series of aforementioned computed residual values. Again the residual values obtained by estimating auto-regressive scheme are to be computed and these are to be added to their corresponding initially derived predicted values to construct a new series. On this newly constructed series again the log - quadratic trend equation is to be estimated by following the same methods as mentioned earlier to obtain the final measure of the growth rate free from either auto-correlation problem or multicollinearity problem or any other econometric problems (like biased growth rate, etc.). To estimate whole period’s growth rates of all the selected performance indicators of life insurance business during 1962-63 to 1999-2000, all the above mentioned procedures have been applied in the study.

4.5.2. Method of Measuring Sub-Period’s Growth Rates

In the growth measurement exercise for performance appraisal, another subject of concern among the scholars is to measure sub-period’s growth rates for the comparison of period-wise (dividing the whole period into a few sub-periods) performance of the variable concerned. In our study also, as mentioned earlier (in Section 4.1), the whole period has initially been divided into two sub-periods (namely, before 1980-81 and from 1980-81 to 1999-2000) and again be divided into four decades (viz., 1960s, 1970s, 1980s and 1990s) to
measure sub-period’s growth rates and to compare the performance of different insurance indicators over different above-mentioned sub-periods.

To measure sub-periods’ growth rates there are three approaches. One of them, used by the early scholars [Das (1978), Rath (1980), etc.], is to fit separate growth function (specially the exponential trend as the span of the sub-period is small) for each sub-period. As this method involves estimation of relatively large number of parameters, it results in the unnecessary loss of degrees of freedom. An alternative approach, most commonly used by the scholars, is to fit a single trend equation with dummy variable. For instance, the form of exponential trend equation with dummy variable for two sub-periods is \( \ln y_t = a + b_1 t + b_2 (t \cdot D) \) where \( D = 0 \) for the first sub-period and \( D = 1 \) for the second sub-period. Here the respective growth rates in the first and second sub-periods are \( b_1 \) and \( (b_1 + b_2) \).

But Boyce (1987) has shown that this approach, along with the first one, for the measurement and comparison of sub-periods’ growth rates suffers from ‘discontinuity bias’ which implies that each of the estimated sub-periods’ growth rates is higher or lower than the estimated growth rate for the whole period. To overcome this difficulty, as per the recommendation of Boyce, we shall estimate ‘Kinked Exponential Model’ of the following form for measuring sub-period’s growth rates:

\[
\ln y_t = a_1 + b_1 \left( D_1 t + \sum_{j=2}^{m} D_j K_i \right) + b_2 \left( D_2 t - \sum_{j=2}^{m} D_j K_1 + \sum_{j=3}^{m} D_j K_2 \right) + \ldots.
\]

\[
+ b_3 \left( D_3 t - \sum_{j=3}^{m} D_j K_{i-1} + \sum_{j=i+1}^{m} D_j K_i \right) + \ldots + b_m \left( D_m t - D_m K_{m-1} \right),
\]

where \( a_1 \) and \( b_i \)'s are the parameters; \( K_i \)s are kinked points and \( D_i \) is the dummy variable for the sub-period \( j \), such that \( D_i = 1 \) for \( j \) th sub-period and \( D_i = 0 \) for sub-periods other than \( j \). Here \( i & j = 1, 2, \ldots, m \).
If the origin is shifted to break point (i.e., to the kinked point), the simplified form of the kinked exponential trend for two sub-periods would be as follows:

$$\ln y_i = a_i + b_1(D_i t) + b_2(D_i t)$$

Where $D_i=1$ for $j$ th sub-period and $D_i=0$ for the other sub-period (where $j=1,2$). Here $b_1$ denotes the estimated growth rate of the first sub-period and $b_2$ denotes the estimated growth rate of the second sub-period. Similarly for four sub-periods (i.e., for four decades), the simplified form of the kinked exponential trend to be estimated in our study is given below:

$$\ln y_i = a_1 + b_1(D_1 t + D_2 K_1 + D_3 K_1 + D_4 K_1) + b_2(D_2 t - D_2 K_1 - D_3 K_1 - D_4 K_1 + D_3 K_2 + D_4 K_2) + b_3(D_3 t - D_3 K_2 + D_4 K_3) + b_4(D_4 t - D_4 K_3)$$

where $b_j$ denotes the estimated growth rates of $j$ th sub-period ($j=1,2,3,4$) and all other notations have their usual meanings as pointed out earlier. It should be mentioned in this connection that, like whole period’s growth rate, measuring sub-periods’ growth rates also through kinked exponential trend equation all the statistical and econometric problems, as pointed out above, are taken care in the study through necessary adjustments. Actually our kinked exponential model gets support from the general class of econometric models with join points (Judge et al. (1980)]

In addition to the measurement of sub-periods’ growth rates, we have enquired whether the estimated growth rates for two sub-periods are statistically significantly different or not. To determine this sort of statistical significance of the trend break (difference between two growth rates) for two sub-periods, i.e., 1962-63 to 1979-80 and 1980-81 to 1999-2000, the following trend equation has been estimated:

$$\ln y_i = a + bt + \beta D_2 t$$

Where dummy variable $D_2=1$ for the second sub-period and $D_2=0$ for the first sub-period; ‘$t$’ denotes the time for the whole period after
normalisation (i.e., shifting origin to the kinked point, mid of 1980) and the estimation of $\beta$ denotes the trend break between first and second sub-period.

4.5.3 Methods of Measuring Stability

As mentioned earlier, along with high growth and level, it is desirable that a performance indicator should be stable (i.e., should behave consistently) over time. The stability of a variable over time is generally measured in terms of its counterpart instability. Actually measurement of instability represents the extent as well as the nature of fluctuations of the concerned variable over a specific time period and this measure of instability has mainly been evolved as a related issue to the growth measurement exercise for evaluating the performance.

In regard to the measurement of instability we get broadly two techniques: (i) summary measure and (ii) trend measure. The trend measure of instability implies the estimation of an appropriate trend (generally the liner trend) on the series of detrend statistic. The detrend statistic implies the value of a variable excluding its trend components. For instance, if a variable (say $y_t$) follows parabolic trend, then on the initial series of $y_t$, the parabolic trend equation (i.e., $y_t = a + bt + ct^2$) is to be estimated and the predicted values (denoted by $\hat{y}_t$) are to be computed. Then the detrend statistic in this situation is represented by either absolute value of $\left| \frac{y_t - \hat{y}_t}{\hat{y}_t} \right|$ or $\left( \frac{y_t - \hat{y}_t}{\hat{y}_t} \right)^2$. After the computation of a series of detrend statistic in the above way if a linear trend is estimated on it, then the estimated coefficient related to time represents the measure of instability. More specifically, if this coefficient is statistically significantly positive (negative), it indicates that the instability of the concerned variable is increasing (decreasing) over time. Otherwise (in case of insignificant coefficient), the variable is stable. But this trend measure of instability suffers from a number of limitations as this measure depends on the arbitrary choice of initial trend as well as of trend estimated on detrend.
statistic and as there are different forms of detrend statistic by which different sorts of variation (namely, seasonal, cyclical and irregular variation) are captured differently. Apart from the existence of these specification errors, only the nature of instability not the magnitude of instability can be deciphered from its trend measure. That’s why, it is better to rely on the summary measure of instability in which there is no specification error and by which the magnitude of instability can be measured.

From the existing studies of Hazell(1982), Dhawan(1983), Hanumantha Rao et al.(1988), etc., it is observed that the prevalent summary measures of instability are standard deviation, variance, standard error of the estimated growth rate, coefficient of variation, etc. Among these summary measures, in our study we have selected the measure of coefficient of variation as this being a relative measure is comparable over different time periods and over different indicators. The actual formula, which has been used in the study as the summary measure of instability, is given below:

Coefficient of variation (CV) = (Standard deviation + Arithmetic mean) x 100

This coefficient of variation is measured for all the performance indicators over the whole period and also for different sub-periods to get the idea on the magnitude as well as the change of instability of the performance indicators over time.

4.6 Methods of Ratio Analysis

Apart from the evaluation of performance of LIC in terms of growth and stability analysis of its few selected performance indicators, the conventional ratio analysis has also been applied for performance appraisal. The ratios which have been taken into consideration for the assessment of the performance of LIC over time may broadly be grouped under two headings: (a) globally recognised ratios specific to the life insurance sector and (b) conventional accounting ratios. For the sake of convenience of the analysis and also to maintain the feature of homogeneity, these ratios have actually
been divided in the study into four groups, namely, (i) Externally measured performance ratios, (ii) internally measured performance ratios (iii) ratios related to the contributions of insurance personnel and (iv) expense ratios. These ratios have been computed by using the raw data obtained from different official publications and the exact formula for computation of these ratios are pointed out one by one below:

Life insurance carrying capacity (LICC) = \[
\frac{\text{Total life insurance in-force in India(SA)}}{\text{GDP}} \times 100.
\]

Premium insurance ratio (PIR) = \[
\frac{\text{Total premium income}}{\text{GDS}} \times 100.
\]

Premium insurance ratio corresponding to household savings (PIRH) = \[
\frac{\text{Total premium income}}{\text{Household savings}} \times 100.
\]

Life insurance penetration (LP) = \[
\frac{\text{No. of policy in individual business + No. of lives in group business}}{\text{Total population}}.
\]

Life insurance density (LID) = \[
\frac{\text{Total premium income}}{\text{GDP}} \times 100.
\]

Life insurance per capita (LIPC) = \[
\frac{\text{Total in-force life insurance business/ price index}}{\text{Total population}} \times 100.
\]

Average rate of return from investment (ARI) = \[
\frac{\text{Return from investment}}{\text{Total investment}} \times 100.
\]

Surplus to total business ratio (STBR) = \[
\frac{\text{Surplus}}{\text{Total sum assured in in-force business}} \times 100.
\]

Life fund as a percentage of total business in-force (SA) (LFTBISA) = \[
\frac{\text{Life fund}}{\text{Total business in-force (SA)}} \times 100.
\]

Percentage of new rural business to total new business (PRTSA) = \[
\frac{\text{Sum assured in new rural business}}{\text{Total sum assured in new business}} \times 100.
\]

Percentage of claim settlement (PCS) = \[
\frac{\text{Claim settled}}{\text{Claim intimated + Claim outstanding}} \times 100.
\]

Total investment to life fund and premium income (TILFPI) = \[
\frac{\text{Total investment}}{\text{(life fund + premium income)}} \times 100.
\]
Ratio of premium income of individual in-force business (ACPH)=Premium income from in-force individual business/Sum assured in individual in-force business.

Average size of insurance in total in-force individual business (AITIB) = Sum assured of individual life insurance policies in-force / Total no. of individual life insurance policies in-force.

Employment contribution ratio (EC)=Employment by LICI/Total employment.

Insurance business per employee (IBPE)=Business in-force (SA)/Total no. of employee.

Insurance business per active agent (IBAA)=Business in-force (SA)/Total no. of active agents.

Insurance business per rupee of employee expenditure (IBREEE)=Total business in-force (SA)/ Total employee expenditure.

Insurance business per officer (IBPO)=Total business in-force (SA)/Total no. of officers other than D0.

Insurance business per development officer (IBPD)=Total business in-force (SA)/Total no. of development officers.

New business expense ratio (NBER)=(Expenses of new business/First year premium income ) X 100.

Renewal expense ratio(RER)=(Renewal expenses/Renewal premium income) X 100.

Overall expenses ratio (OER)=(Total management expenses/Total premium income) X 100.

Through all these ratios, the performance of the life insurance business in India has been examined for an in-depth analysis. To be more specific, all these selected performance ratios have been presented first graphically through line diagrams (measuring the concerned ratio on the vertical axis and the time on the horizontal axis) to get an overall idea on the movements of these ratios over time. Along with this graphical analysis, we have also
applied the statistical techniques of 'mean difference test' to examine whether the level of a particular ratio has significantly changed over different sub-periods or not.

For mean difference test, sub-period-wise (for two sub-periods and also for four decades) means of each ratio have been calculated. Let, \( \bar{r}_i \) be the mean of a ratio in \( i \)th sub-period and \( \bar{r}_j \) be the mean of that ratio in \( j \)th sub-period. Then the test statistic for mean difference or joint mean test is:

\[
\frac{(\bar{r}_j - \bar{r}_i) - (\mu_j - \mu_i)}{S.E(\bar{r}_j - \bar{r}_i)}
\]

Where S.E. denotes standard error and the symbol \( \mu \) represents population mean. Here the \( i \)th sub-period precedes the \( j \)th sub-period (\( i \neq j = 1,2 \) or \( =1,2,3,4 \)). For this test the null hypothesis is \( H_0: \mu_i = \mu_j \) and the alternative hypothesis is \( H_1: \mu_i \neq \mu_j \). So for the mean difference test we shall apply two tailed test and under null hypothesis, the test statistic is supposed to follow Student's t distribution. From this mean difference test we shall be able to examine statistically whether particular ratio significantly increases or decreases or remains unchanged (in case of insignificant mean difference) between two adjacent sub-periods. That is, how far the performance of LICI measured in terms of selected ratios has changed periodically, can be examined from the analysis of joint mean test of the ratios.

Along with the overall examination of the graphical movements of the ratios and carrying out the mean difference test of the ratios, to know the magnitude of their changes over time one sophisticated statistical analysis has also been applied in this area. To avoid the problem of multicollinearity, to increase the degrees of freedom in the statistical test and to club the ratios belonging to the same group, first, 'principal component analysis' based on Kaiser's criterion [Koutsoyiannis (1977) Pr. 424-436] has been applied. Kaiser's criterion implies that only those components will be extracted whose Eigen
value is greater than one. Along with the basic estimated results of principal component analysis (or alternatively known as factor analysis), the results of 'Bartlett's test of sphericity' and component matrix are also presented. Bartlett's test of sphericity tests whether the correlation matrix of the variables to be clubbed by principal component analysis is an identity matrix, which would, in turn indicate that the model of factor analysis is inappropriate. In other words, a statistically significant result of Bartlett's test of sphericity establishes that principal component or factor analysis is not an inappropriate exercise. The figures of component matrix, on the other hand, represent the proportional contributions of the original variables in the constructed principal components. It should be mentioned in this connection that the principal component has been constructed after transforming the original explanatory variables in their standardised forms [e.g., for explanatory variable $x_i$, the standardised form is $(x_i - E(x_i))/\sigma x_i$].

Principal component analysis has been applied to club four sets of more or less homogeneous ratios in the form of four constructed respective factors (or principal components). These four sets of ratios have been termed as (i) Efficiency ratio, (ii) Productivity ratio, (iii) Income ratio and (iv) Expense ratio. Within efficiency ratio, life insurance density (LID), life insurance carrying capacity (LICC), premium insurance ratio (PIR), and life insurance per capita (LIPC) are taken into consideration. Similarly, the constructed principal component for productivity ratio includes the basic ratios like, insurance business per employee (IBPE), insurance business per active agent (IBAA), insurance business per rupee of employee expenditure (IBREEE), insurance business per development officer (IBPD) and insurance business per officer (IBPO). Principal component for income ratio is constructed to club only two ratios, namely, average rate of return on investment (ARI) and surplus to total business ratio (STBR). Similarly, two ratios, viz., overall expenses ratio (OER)
and renewal expense ratio (RER) have been clubbed to construct the principal component for expense ratio.

All these four constructed principal components for four sets of ratios have been used as the explanatory variables for the regression analysis in which the explained or the dependent variable is the dummy variable. This technique of regression analysis has clearly been explained by Koutsoyiannis (1977, Pp. 283-84). In our case, the regression equation with dummy dependent variable will be as follows:

\[ D_t = b_0 + b_1 P^1_t + b_2 P^2_t + b_3 P^3_t + b_4 P^4_t + u_t \]

Where \( D_t \) denotes dummy dependent variable and \( u_t \) is the disturbance term. \( b_j \)s (j=0,1,2,3,4) are the regression parameters to be estimated and \( P^i_t \) denotes the constructed principal component for ith set of ratios (i=1,2,3,4) at time t.

Actually we have estimated the above mentioned type of regression equation twice for two types of dummy dependent variable (namely, \( D_{1t} \) and \( D_{2t} \)) to examine the effect of reforms on the performance ratios of LICCI prior to (i) 1980-81 and (ii) 1992-93. More specifically, these two dummy dependent variables for two estimated regression equations are defined as follows:

\[ D_{1t} = 0, \text{ when } t \leq 1980-81 \]
\[ = 1, \text{ when } t > 1980-81. \]

And \( D_{2t} = 0, \text{ when } t \leq 1992-93 \]
\[ = 1, \text{ when } t > 1992-93. \]

From the estimated results of the regression equations with dummy dependent variable we shall be able to know whether the selected ratios (which have been taken into consideration for the construction of the principal components) significantly change during the reforms period in comparison to their values in the pre-reforms period.
4.7 Methods for Identifying Socio-Economic Determinants of Life Insurance Business

For identifying the socio-economic factors on which the performance of the life insurance in India depends, multiple regression equations have been estimated both over time (for each of the thirteen major states of India and also for the country as a whole) and cross sectionally (over the major states of India on three time points namely, 1970-71, 1980-81 and 1990-91). The analyses have been organized in the above mentioned way as per the availability of the necessary data, which are explained as follows:

To enquire how far the general economic well being affects the life insurance business in the country, as per the availability of data, multiple regression equation of the following form has been estimated for the thirteen selected major states (namely, Andhra Pradesh, Assam, Bihar, Gujarat, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Orissa, Rajasthan, Tamil Nadu, Uttar Pradesh and West Bengal) of India for the period 1970-71 to 1992-93:

Total sum assured of life insurance business in the state = \( f(\text{net state domestic product, estimated population of the state}) \).

But from this type of estimated regression equation, most probably spurious results will crop up because the dependent variable and also the above mentioned two independent variables are subject to the increasing function of time. Therefore, to obtain the actual relationship between the dependent and the independent variables, trend component has to be excluded from each of the three above-mentioned variables by estimating linear trend equation \( y_t = a + bt \). That is, subtracting the predicted values (obtained from the estimated linear trend) from the corresponding original values of each of the three variables, multiple regression equation of the same above-mentioned form has again been estimated. This technique of re-estimation of multiple regression equation will help us to exclude trend...
with the presence of the effect of inflation, if any, from the variables and to decipher their true inter-relationships.

But the estimation of the above type of multiple regression equation for each of the major states using time series data suffers from one major limitation. The limitation is the non-inclusion of some important explanatory variables in the regression equation; that leads to the estimation of biased results econometrically. To avoid this sort of limitation, some more independent variables should be included. But the time series data for no more related independent variables are available for the states in India. To accommodate few more important independent variables (as per the availability of data) the following multiple regression equation has been estimated cross sectionally (i.e., over thirteen major states in India) for each of the three time points (viz, 1970-71, 1980-81 and 1990-91):

\[ Y_{ij} = a_0 + a_1 x_{ij} + a_2 x_{2j} + a_3 x_{3j} + a_4 x_{4j} + a_5 x_{5j} + a_6 x_{6j} \]

Where \( Y_{ij} \) denotes the total sum assured of life insurance business in jth state. \( x_{ij} \) denotes net domestic product of jth state. Similarly, \( x_{2j} \)=literacy ratio in jth state, \( x_{3j} \)=average size of household in jth state, \( x_{4j} \)=proportion of male population in jth state, \( x_{5j} \)=proportion of urban population in jth state and \( x_{6j} \)=proportion of able bodied persons in total population in jth state \([j=1, 2, .., 13]\).

Estimating multiple regression equation of the above form cross sectionally for three time points separately, we shall be able to enquire the relative importance of the socio-economic variables in determining life insurance business in India and also to know the changes of their relative importance over different time points, if any. But in this cross sectional regression analysis the main limitation is that due to non availability of data the analysis could not be extended beyond the time point 1990-91.

To extend the period starting from 1962-63 to 1999-2000 (i.e., the whole period of our study), multiple regression equation in determining socio-economic determinants of life insurance business in India has been estimated
using the time series data for the country as a whole. Time series regression analysis is also required to corroborate or to supplement the results obtained from cross sectional regression analysis. Moreover, with this time series analysis few more additional explanatory variables have been included (altogether eleven in number) along with the inclusion of dummy explanatory variable by which the effect of a specific period on life insurance business can be examined. More specifically, in regard to the dummy independent variable the same multiple regression equation has been estimated separately twice taking dummy variable \( D_1 \) (defined as \( D_1 = 0 \) for the years prior to 1980-81, otherwise \( D_1 = 1 \)) at one time and the dummy variable \( D_2 \) (defined as \( D_2 = 0 \) when time <1991-92 and equal to 1 when time \( \geq 1991-92 \)) for the other; with these two dummy variables we shall be able to enquire the influences of four sub-periods on life insurance business [namely, (i) the period prior to the introduction of reforms suggested by the Era Sezhiyan Committee, (ii) the period after the introduction of the reforms suggested by the Era Sezhiyan Committee, (iii) the period prior to adopting general new economic reforms and (iv) during the new economic reforms period].

With the introduction of large number of explanatory variables in relation to the number of years (i.e., number of observations) in a time series regression analysis two undesirable features will crop up: (i) this will lead to the reduction of the degrees of freedom and the robustness in estimation and (ii) some of the explanatory variables may be related among themselves leading to the problem of multicollinearity. To avoid these two problems and also to capture the advantage of the inclusion of more explanatory variables we have applied principal component or factor analysis before the estimation of time series multiple regression analysis. The detailed procedure of principal component analysis has already been illustrated in Section 4.6 of this chapter. With the help of principal component analysis here, all the explanatory variables related to demographic factors, like, proportion of male persons,
proportion of able-bodied persons and proportion of literate persons (corresponding to total population) have been clubbed and termed as \( \text{Fac}_2 \) that represents the joint effect of the demographic factors. Therefore, for time series analysis for the country as a whole, the form of the multiple regression equation to be estimated for determining socio-economic determinant of life insurance business would be as follows:

\[
Y_t = a + b_0 \text{Fac}_1 + b_1 x_{1t} + b_2 x_{2t} + b_3 x_{3t} + b_4 x_{4t} + b_5 x_{5t} + b_6 x_{6t} + b_7 D_1 \text{ or } D_2
\]

Where \( a \) and \( b_0 \) are the regression coefficients to be estimated, \( \text{Fac}_1 \) and Dummy variable \( D_1 \) or \( D_2 \) have already been defined, \( y_t \) denotes the dependent variable in the year \( t \) (\( t=1962-63, 1963-64 \ldots . \ 1999-2000 \)) and

\( x_{1t} = \text{Real gross domestic savings in } t \text{th year}, \)

\( x_{2t} = \text{Proportion of population engaged in public and organized sectors in } t \text{th year}, \)

\( x_{3t} = \text{Average size of household in } t \text{th year}, \)

\( x_{4t} = \text{No. of persons engaged in marketing insurance products in } t\text{'th year}, \)

\( x_{5t} = \text{Average premium rate in the year } t \)

and \( x_{6t} = \text{Relative rate of return from LICI (i.e., average bonus rate offered by LICI divided by the average interest rate of term deposits in the commercial banks).} \)

In this type of time series analysis for the country as a whole we are also able to accommodate different forms of dependent variables (viz., total real sum assured in life insurance business, life insurance penetration, life insurance density and life insurance carrying capacity) as different measures of life insurance business in India. So actually four types of multiple regression equation have been estimated over time taking four types of the dependent variable, \( y_t \).

From the estimated regression results (to be illustrated in Chapter VII), it is observed that bonus rate (i.e., the relative rate of return form LICI) plays a very important role in determining the expansion of life insurance business.
To throw more light on this particular aspect, we have enquired through time series regression analysis whether bonus rate statistically significantly depends on average rate of return from investment of LIC and also whether average rate of return from investment of LIC statistically significantly depends on the share of investment in private sector. In all these regression equations also the effect of the periodic changes due to the introduction of economic reforms has been examined by introducing dummy independent variable. Lastly, we have examined the presence of popular ‘March Phenomenon’ in life insurance business in India. That is, to know whether life insurance policies are considered as the tax saving weapon, we have also regressed new insurance business during the months from January to March of each year on the total income tax paid by the individuals and Hindu Undivided Families (who actually get tax benefit by purchasing the policies of LIC) in India. It should be mentioned in this connection that the significant positive relationship between new insurance business and income tax paid (as defined above) implies the existence of ‘March Phenomenon’.