CHAPTER 3
METHOD AND MATERIAL

3.1 Introduction

The chapter is organized in five sections. The second section 3.2 provides the details of different sources of data used in the study. The list of the variables to be used in the study is discussed in section 3.3. A brief discussion on the methods adopted to workout composite indices of educational development, measuring growth rates and relationship between variables is provided in section 3.4. The measurement of impact of education on economic development has been discussed in the final section. This section, also presents the details of Granger’s Causality Test which is employed to determine the causal relationship between public expenditure on education, education development and economic growth.

3.2 Nature And Sources of Data

The present study is based on the Himachal Pradesh’s official data. The whole analysis is primarily based on secondary sources of data and the required data is collected from various published sources by the government of India and other official agencies. The specific publications that have been referred in the present study are mentioned below:

- Education in India, vol.1, Ministry of Education and Social welfare, Government of India.
- Finance Accounts, Controller and Editor General, Government of Himachal Pradesh.
- Economics Survey of India, Ministry of India, New Delhi.
3.2.1 Period of Analysis

Ideally, it would have been worth while to undertake the present study since the formation of Himachal Pradesh in 1966. However, it was with the reorganization of districts in 1971, that the state came into its present form. To estimate the growth of education in the state, the study broadly covers the period of 34 years, from 1971-1972 to 2004-2005. However, for the analysis of public expenditure on education and impact of education on economic development, the period covered is from 1980-1981 to 2004-2005 due to the non-availability of data pertaining to Net State Domestic Product at 1993-1994 prices for earlier period.

3.3 Methodology and Technique of Analysis

Besides using simple statistical devices like the growth rates, coefficient of correlation and coefficient of variation in the study, method of Principal Component Analysis (PCA) has also been utilized to work out composite indices of educational development. The relationship between expenditure on education, education
development and economic development, has been evaluated by employing Granger's Causality Test.

3.3.1 Methodology for Computing Growth Rates

The exponential growth rates are calculated by using the following types of functions.

\[ y_i = a (b_i)^t \]

According to which

\[ \log y_i = \log a + t \log b_i \]

i.e. \( \log y_i \) is a linear function of \( t \).

Annual compound growth rate will be computed as:

\[ r = [(\text{antilog } b - 1) \times 100] = \text{compound growth rate (in percent)} \]

3.3.2 Standard Deviation

To analyse the variation in the indicators of public expenditure on education and educational attainment, standard deviation (\( \sigma \)) and coefficient of variation would be estimated by

\[ \sigma = \sqrt{\frac{\sum x^2}{N}} \]
(where 'x’ refers to deviation of items from their means, ‘Σ’ denotes summation and ‘N’ denotes number of observations).

\[
C.V = \frac{\sigma}{\bar{X}} \times 100
\]

### 3.3.3 Coefficient of Correlation

To find linkages among the determinants of education and economic development Karl Pearson’s correlation co-efficient \( r \) would be computed by:

\[
r = \frac{\Sigma X \times Y}{\sqrt{\Sigma x^2 \times \Sigma y^2}}
\]

(\( x \) and \( y \) refers to deviation of items from their means, ‘Σ’ denotes summation)

### 3.4 Principal Component Analysis

The use of composite index to examine interstate or inter-district variation in educational development is very common (Tilak 1982). Several methods have been adopted by different authors to calculate the composite index. For instance Rudolph and Rudolph (1969), Pandit (1977), Heyneman (1979) used simple ranking method. Panchamukhi (1970) and Tilak (1981) applied principal component analysis and Reddy (1977) used Taxonomic Method. Tilak (1982) also constructed composite index by using cost proportion as the weights. Rao (1986), following the methods of principal component analysis, identified the backward states taking into account 15 indicators of different socio-economic characteristics.
All the methods of composite index, adopted by various researchers have their own merits and demerits. The principal component analysis is preferred over other methods when the variables used for the estimation of composite index are in different units of measurement i.e., enrolment of students, number of educational institutions, per capita income in rupees etc. The advantages of the Principal Component Analysis are that it helps in reducing the dimension of a complex multivariate problem and it takes the correlation matrix into account and produces components which are uncorrelated with one another and thus, helps to overcome the problem of multi-collinearity. Since our study has aimed at estimating the composite indices of education and economic development, both having varied units of measurement, we would prefer to use this method in our study over others. Also this method is simple as well as logical in its approach.

3.4.1 Selection of Indicators

In order to have a clear picture of educational development over time, on the basis of various disaggregated variables, deflated as far as possible for changes in relevant population, composite indices of quantitative, qualitative and over all development of education are constructed on the basis of the first principal component.

Various studies on education have used different indicators as proxies for education. The main reasons for this, other than the difference in conceptualizing education, has been the lack of reliable data, especially in developing countries. In the present study, the following quantitative and qualitative indicators have been incorporated as proxy for education development.
A. Quantitative Indicators

Nine variables indicative of the quantitative aspects of educational development are taken up. These are:

1. Primary schools per lakh of population (X₁)
2. Middle schools per lakh of population (X₂)
3. High and Higher Secondary schools per lakh of population (X₃)
4. Higher level institutions per lakh of population (X₄)
5. Total number of educational institution per 100 square kilometer (X₅)
6. Enrolment-ratio at the primary level defined as percentage of Enrolment in classes 1 to 5 to population of age group 6-11 years (X₆)
7. Enrolment ratio at the middle level defined as the percentage Enrolment in classes VI to VIII to population of age group 11-14 years (X₇)
8. Enrolment ratio at the high and higher secondary level defined as the percentage of enrolment in classes IX to XII to the population of age group 14-18 years (X₈)
9. Enrolment ratio at the higher level defined as the percentage of enrolment in higher classes to the population of age group 18-24 years (X₉)

B. Qualitative Indicators

Seven variables indicative of the qualitative aspect of educational development are taken into account. These are:

1. The percentage of trained teachers to total number of Teachers in primary level (X₁₀).
2. The percentage of trained teachers to total number of teachers in middle level (X₁₁).
3. The percentage of trained teachers to total number of teachers in high and higher secondary level (X₁₂).
4. Teacher student ratio, at primary level ($X_{13}$).
5. Teacher student ratio, at middle level ($X_{14}$).
6. Teacher students ratio, at high and higher secondary level ($X_{15}$).
7. Teacher student ratio, at higher level ($X_{16}$).

3.4.2. Technique of Principal Component Analysis

The principal component method first developed by Hotelling (1933) has been adopted for computation of composite indices of quantitative, qualitative and overall educational development in Himachal Pradesh over time. Principal components can be calculated from the variance-covariance matrix of the variables, if we are dealing with the variables in their original form. They can also be evaluated from the correlation matrix between the variables if we are dealing with the variables in their original form. They can also be evaluated from the correlation matrix between variables if we are dealing with standardized values, measured as deviations of the variables from their means and divided by the standard deviations. In our analysis, we concern ourselves with the standardized variables and with the calculation of the principal components from the correlation matrix.

The principal component analysis enables us to determine a vector known as the principal component, linearly dependent on the standardized constituent variables and having the maximum sum of squared correlation with the variables. There can be as many principal components as there are variables. The aim of the method of principal components is the construction of new variables $P$ called 'Principal Components' which are linear combination of $X$'s such that
\[ P_1 = a_{11} X_1 + a_{12} X_2 + \ldots + a_{1k} X_k \]
\[ \vdots \]
\[ P_k = a_{k1} X_1 + a_{k2} X_2 + \ldots + a_{kk} X_k \]

The \( a_{ij} \)'s called loadings are chosen so that the constructed principal components satisfy two conditions: (a) the principal components are uncorrelated and (b) the first principal component \( P_1 \) absorbs and accounts for the maximum possible proportion of the variation, in the set of all \( X \)'s, the second principal component absorbs the maximum of the remaining variation and so on.

The variables \( X_j \) are first standardized using the simple formula

\[ Z_j = (X_j - \bar{X}) / \sigma_x \]

where \( \sigma_x \) is the standard deviation of \( X_j \)

With standardized values \( z_i \), then the first principal component \( P_1 \) can be written as

\[ P_1 = a_{11} z_1 + a_{12} z_2 + \ldots + a_{1n} z_n, \]

where

\[ a_{11} = \frac{\sum_{i=1}^{n} r_{X_i X_j}}{\sqrt{\sum_{j=1}^{n} \sum_{i=1}^{n} r_{X_i X_j}}} \]

with \( r \) indicating the coefficient of correlation. The coefficients of \( z_i \) in the equation giving the first principal component are referred to
as loadings and are also denoted by $l_{1i}$ with $\lambda_1$ the latent root of the first principal component defined as

$$\lambda_1 = \sum_{i=1}^{n} l_{1i}$$

The percentage contribution of $P_1$ in the total variance in the standardized variables is defined by

$$\frac{\lambda_1}{n} \times 100$$

The Second Principal Component is computed from a residual correlation matrix, obtained by deducting $a_{1i} a_{ij}$ from $r_{xi} x_j$ in each cell $(i, j)$ in the original correlation matrix, with subsequent principal components being computed from residual correlation matrices, similarly obtained at each stage.

### 3.4.3 Limitation of Principal Component Analysis

The general limitations of principal component analysis are well known and have been documented in detail in most books (Koutsoyiannis, 1977). One is that the assumption about the variable being linearly related, may not always hold true. Secondly, there is no objective way of deciding whether we should use the original variables or the standardized ones and the values of the principal components vary depending upon the value of the variables used. Thirdly, as pointed out by many, the practical interpretation of these principal components, which are artificial orthogonal variables, is an important research problem. Further, there are no universally accepted rules regarding the testing of the significance of the
loadings. There is also no unambiguous scientific basis regarding the number of principal components to be taken nor can one combine the different principal components to form a composite index.

3.4.4 Nature of Present Application

In the present study the variables of educational development are grouped into two, which are indicative of quantitative and qualitative development respectively - with the first principal component being worked out for each group separately. In the light of the findings by Kundu (1980) and Pal (1972) that the first principal component of the two meaningful groups of variables can be treated as two distinct variables and the first principal component of these, worked out to evolve a realistic summary measure for both the groups of variables together, over-all educational development is sought to be obtained by working out the first principal component of the indicators of qualitative and quantitative development.

A particular issue related to the application of the principal component analysis in this study, is worth mentioning. In regional studies the principal component analysis is often used to indicate the level of development of a region on the basis of the values of the variables in the different regions at a particular of time (Hagood and Price, 1952). The weighting pattern of the principal component is thus the same for the different regions, with the values of the principal component differing due to the differences in the values of the variables in the different regions. However, Schilderinck (1970) and also Sardamoni (1981), in fact used principal component analysis to compare the behaviour of the variables in different periods and to investigate the influence of some variables upon some others over time, the approach here seems warranted though subjected to limitations.
3.5 Measuring Impact of Education on Economic Development

Various methods have been adopted by different researchers to assess the contribution of education to economic development. The important methods are:
1. Residual method,
2. Growth accounting equation method,
3. Simple correlation method and
4. Econometric method.

1. Residual Method

Using the residual method, the contribution of education has been estimated considering two factors of production, namely labour and capital, with the help of the linear homogeneous Cobb-Douglas production function. The portion of the increased output, which could not be accounted for by these two factors, has been termed as residual or unexplained third factor, which was supposed to be comprised of technological progress, economies of scale, external economies, improved health, education and skill of the labour force, better management, changes in the product mix etc. Education is regarded as important factor out of all. But the residual method has its own limitations. Aggregating output, capital and labour over a long period is rather very difficult and the composition of the residual factor is some what ambiguous.

2. Growth Accounting Equation Method

This method is just an extension of the residual approach which can be easily extended to include human capital as a
determinant of the economy's growth rate based on a dynamic Cobb-Douglas aggregate production function. To reduce the residue, researchers have started incorporating education and other variables in the aggregate production function. Whether the level of income influences the level of education or it's the other way round still remained unanswered.

3. **Correlation Method**

Correlation method observes the relationship between the indicators of education development either at a point of time or over a period of time. Correlation approach has its own drawbacks. It neither brings out the cause and effect of relationship nor tells about the magnitude of educational contribution to economic growth.

4. **Econometric Method**

In order to identify the cause and effect relationship between education and economic development, lately, attempts have been made to relate inputs to outputs with the help of econometric tools like simultaneous equation model, ordinary least square methods etc.

3.5.1 **Tests of Causality**

**The Granger Test**

In order to investigate the impact of education on Himachal Pradesh's economic development, the present study adopts a time series based empirical strategy, Unit Root Tests would be followed by the Granger Causality Test.

We employ data for Himachal Pradesh, covering the period from 1980–81 to 2004–2005. Data are annual and made available from various published and unpublished secondary sources.
mentioned in detail in section 3.2 of this Chapter. The Per Capita NSDP (in national currency and at constant prices) and its relative growth rate has been used as a proxy of economic development while for educational development proxies, we have used composite index of economic development and per capita public expenditure on education.

For determining the causal relationship between education and economic growth, the following hypothesis is tested:

\[
H_0 : E \left( \Delta y_t / \Delta y_{t-1}, \Delta y_{t-2}, \ldots, \Delta y_{t-n}, \right.
\Delta x_{t-1}, \Delta x_{t-2}, \ldots, \Delta x_{t-m}, \Delta z_{t-1},
\Delta z_{t-2}, \ldots, \Delta z_{t-m} \right) = E \left( \Delta y_t / \Delta y_{t-1}, \right.
\Delta y_{t-2}, \ldots, \Delta y_{t-n}, \Delta z_{t-1},
\Delta z_{t-2}, \ldots, \Delta z_{t-m} \right)
\]

for all \( m > 0 \) - No Granger Causality. Against the alternative of,

\[
H_1 : E \left( \Delta y_t / \Delta y_{t-1}, \Delta y_{t-2}, \ldots, \Delta y_{t-n}, \right.
\Delta x_{t-1}, \Delta x_{t-2}, \ldots, \Delta x_{t-m}, \Delta z_{t-1},
\Delta z_{t-2}, \ldots, \Delta z_{t-m} \right) \neq E \left( \Delta y_t / \Delta y_{t-1}, \right.
\Delta y_{t-2}, \ldots, \Delta y_{t-n}, \Delta z_{t-1},
\Delta z_{t-2}, \ldots, \Delta z_{t-m} \right)
\]

For some \( m > 0 \) - Granger Causality

Here \( \Delta y \) represent the first difference of the per capita NSDP, \( \Delta x \) represents the first difference of the composite index of educational development, and \( \Delta z \) represents the first difference of the per capita expenditure on education. Before conducting any of
the tests, all the relevant series are tested for stationarity, since
standard inference procedures do not apply to regression which
contain an integrated dependent variable or integrated regressors.
Moreover, in the presence of non-stationary variables, there is a
possibility of spurious regression. A formal method to test for
stationarity of a series is the Unit Root Test. To this effect the
standard Augmented Dicky Fuller (ADF) tests would be utilized.
Next, the following model would be formulated to test for a causal
relation,

\[ \Delta y_t = \delta_0 + \sum_{j=1}^{m_1} \delta_{tj} \Delta y_{t-j} + \sum_{j=1}^{m_2} \delta_{2j} \Delta x_{t-j} + \sum_{j=1}^{m_3} \delta_{3j} \Delta z_{t-j} + e_{t,t} \] ...... (1)

For the lagged variables appearing on the right hand side, the
number of lags is determined using the Akaike Information Criterion
(AIC) and Shwartz Criterion (SC) and the lag that gives the lowest
AIC and SC and best fit is chosen. Adding lagged values of the
dependent variable on the right-hand side, other than fulfilling the
Granger Causality requirement, also reduces or eliminates the
problem of spurious results due to serial correlation. A major part of
the analysis depends on the choice of lag length since the result of
the tests of causality rely heavily on the time lags being imposed. If
\( \delta_{2j} \) and / or \( \delta_{3j} \) are found to be statistically significant and different
from zero, we reject \( H_0 \) and accept \( H_1 \). The hypothesis is tested using
a standard F-test.