Having formulated the problem and provided the rationale thereof, in the preceding chapter, we now turn to its conceptual and statistical basis—selection of unit of analysis, methodology, and sources of data.

STATE AS UNIT OF ANALYSIS—

The administrative unit, State, follows a great history of unification and division within the national boundaries. Major administrative break was brought by the formation of thirteen States and six union territories that came into being on the recommendations of State Reorganization Commission in 1956. Various legislative provisions since then, until recently in 2000, have reorganized these administrative divisions on one or the other pretext.

The thirteen States and six union territories that came into being in 1956 include, Andhra Pradesh, Assam, Bihar, Bombay, Jammu and Kashmir, Kerala, Madhya Pradesh, Madras, Mysore, Orissa, Punjab, Rajasthan and West Bengal among the States, and Andaman-Nicobar Island, Delhi, Himachal Pradesh, Tripura, Lakshadeep, and Amindivi were Islands the union territories. Subsequent years saw further divisions on this base.

As per 2001 Census, India is comprised of 27 States excluding Delhi and six union territories, while there were 25 States and seven union territories of 1991; and 15 States and 12 union territories of 1961. A preliminary survey of literature and data sources brought to the fore major difficulties in aggregating data on various counts across States in the presence of major time-to-time changes in the States boundaries. In order to minimize such difficulties, the present study has been restricted to 17 major States. Due to the non-availability of complete dataset on the selected variables for the development construct, we could not cover the stated period of 1951-2001, instead the
The selected States for the present study include- Andhra Pradesh (AP); Assam (AS); Bihar (BH); Gujarat (GJ); Haryana (HR); Himachal Pradesh (HP); Jammu and Kashmir (JK); Karnataka (KR); Kerala (KE); Madhya Pradesh (MP); Maharashtra (MH); Orissa (OR); Punjab (PB); Rajasthan (RJ); Tamil Nadu (TN); Uttar Pradesh (UP); and West Bengal (WB).
MAIN FOCUS OF THE STUDY

As stated earlier, the study aims at -

- Testing neo-classical convergence (absolute and conditional) hypothesis for per capita net state domestic product and development levels of States during 1971-2001;
- Tracing the effects of spatial dependence of States on their growth prospects and estimating probable time to reach to their respective steady state paths;
- Constructing composite indices of development, based on five broad dimensions of development for examining the levels and trends of development across States, testing for any possible (statistically significant) changes in the levels of development over time;
- Measuring the extent of disparities in the levels of development across States, and the movement there-off, and
- Examining the possibilities of absolute and conditional convergence on the basis of individual sectoral indices as well as the composite index of development.

Alternatively stated, these issues boil down to the following set of hypotheses.

Hypotheses- We hypothesize that:

1. disparities in development have been on the increase in independent India;
2. the increase in inter-State disparities further accentuated in the post reforms period;
3. The States of India do not show β-convergence in their development experience;
4. Plan expenditure mitigated developmental gaps across States; and
5. Inter-State disparities in development would not only persist but would aggravate over time.

Examining each of the above aspects or testing of this hypothesis call for an elaborate statistical procedure. The details of the same are sought to be provided in the following section.
METHODOLOGY

The statistical procedures that are used to (i) measure growth, (ii) treat raw data on variables measured in different units, (iii) assign weights to the transformed variables for meaningful analysis, (iv) construct a composite index with given set of variables, (v) measure the extent of disparities across States, (vi) test absolute and conditional convergence in development across States over time, and (vii) identification of the main sources of development, are discussed here.

Measures of Growth

Rate of change over time measures the proportional change in an entity/variable in a given period over its base year value. The ‘average annual rate of change’ calculated as arithmetic mean of the yearly change may not present the trend and actual rate of change. It is biased towards extreme annual rates. For instance, a random annual series with values, 1375, 1895, 1255 gives an average change of 2.02 per cent per annum. Starting with base value 1375, an 2.02 per cent increase in 1375 add 27.775 to 1375 to give 1402.775 in the subsequent year. Again, 2.02 per cent increase in 1402.775 add 28.336 more to 1402.775 to give 1431.111 in the next year. Thus, growing at an average annual rate of 2.02 per cent, the end value over two years time would give 1431.111, which is far from the actual end value of 1255 of a random series.

Instead, the study makes use of a more sophisticated measure of change, compound annual rate of growth CARG that annualizes the change. The CARG, least square growth rates uniformly distributes change over the years, but ignore fluctuations. We estimate growth rate, g, by fitting a linear regression trend line to the logarithmic annual values of the variable in the relevant period. The logarithmic transformation of the compound growth equation is represented in the regression equation that takes the form-

\[ \ln X_t = a + bt, \]

which is equivalent to

\[ \ln X_t = \ln X_0 + t \ln (1 + r), \]

44
where 'X' is the variable, 't' is time, and \( a = \ln X_0 \) and \( b = \ln (1 + r) \) are parameters to be estimated. If \( b^* \) is the least-squares estimate of \( b \), the average annual growth rate, \( r \), is obtained as \( [\exp(b^*) - 1] \times 100\% \).

Methods to Transform the Dataset

As far as the examination of the levels of development and extent of disparities in single indicators is concerned, the need to transform variables does not arise. A simple look at the measures such as means, standard deviations and coefficient of variation in case of univariate analysis contributes enough to our understanding of the data. In order to construct a composite measure for a multidimensional phenomenon of development we need to put data on different counts in standard units. The need to make the dataset scale free arises from the fact we want study is to examine relative performance of States on different counts, each having a different unit of measurement. Removing scale bias is in itself the first step in various procedures of indexing a multivariate dataset.

From various methods of transforming the dataset, most common is the one followed by the Human Development Reports. The mathematical form used in the Human Development Reports for standardization of variables, represented by \( Z_{ij} \)'s for the \( i^{th} \) State and the \( j^{th} \) takes the following form—

\[
Z_{ij} = \frac{(X_{ij} - \text{min } X_{ij})}{(\text{max } X_{ij} - \text{min } X_{ij})},
\]

where \( X_{ij} \) represent the original values of the \( j^{th} \) indicator for the \( i^{th} \) State, \( i = (1,2,\ldots,N=17) \) and \( j = (1,2,\ldots,k) \). \( \text{min } X_{ij} \) and \( \text{max } X_{ij} \) are respectively the minimum and the maximum of \( (X_{i1}, X_{i2},\ldots,X_{ik}) \). The scaled values, \( Z_{ij} \) varies form zero to one in this form.

\(^1\) The World Bank Reports
Though ‘standardized z-scores’ approach is questionable for equalizing variance and length of all the indicators, we have opted for it for our study for its wide application base in social science research.

The z-transformation to the variables makes variation comparable. The square-root of the average squared mean deviations measures the variation around the mean. Dividing the square of the mean deviations by the square-root of the average squared mean deviation brings back the variables in original units. The subtraction of mean from each entry of a variable does not change the magnitude of variation.

The mathematical form of standardization, represented by $Z_{ij}$'s for the $i^{th}$ State and the $j^{th}$ variable is given below:

$$Z_{ij} = \frac{(X_{ij} - \mu)}{\sigma_i} \quad (2)$$

where $\mu$ is the mean of $X_j$ and $\sigma_i$ is the standard deviation for $X_j$. The z-values vary between $-3$ and $+3$.

For a given set of $k$ variables on $N$ States it takes the following form:

$$\begin{pmatrix}
Z_{11} & Z_{12} & \cdots & Z_{1k} \\
Z_{21} & Z_{22} & \cdots & Z_{2k} \\
\vdots & \vdots & \ddots & \vdots \\
Z_{N1} & Z_{N2} & \cdots & Z_{Nk}
\end{pmatrix} \quad (A)$$

where $Z_{ij}$ is the standardized values of $j^{th}$ variables of $i^{th}$ State, $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, k$.

The z-score approach also seems to satisfy the two axioms (R and S) of transformation proposed by Kundu (2002).²

² Refer Rummel, 1976.
³ Standard deviation is discussed later in this chapter in measures of inequality.
Methods for Assigning Weights

The nature and magnitude of interrelationships among the selected set of variables is itself a good determinant of weights to be assigned to each variable. A good number of studies using different methods to construct a composite index attach great importance to various methods of assigning weights to each variable included in the model.

State relatives, regional economic distances, regression analysis, gini concentration coefficients, and principal component analysis are few such measures that are extensively used to assign weights in the multivariate aggregation models. The beauty of these procedures lies in their ability to retain the original variation that makes comparison between regions dependable.

Gini concentration argue that the more skewed distribution represent more deprived States, and more deprived a State is in particular aspect the more important that aspect is to the State. Thus, it gives more weights to more volatile variables. The gini coefficient of each indicator is used to assign weight to the respective indicator. In mathematical form, it is:

\[ g_j = \frac{2 \text{cov}(X, r_j)}{n \mu}, \]

where \( g_j \) is the gini coefficient of the \( j \)th variable is also the weight to the \( j \)th variable, \( \text{cov}(X, r_j) \) is the covariance of the indicator \( X \) with its ranks for all States. and \( \mu \) is the mean of variable ‘X’.

Another approach to determine weights assumes that weights vary inversely to the variation in the respective indicator and is computed as below:

\[ w_i = \frac{m}{\sqrt{\text{Var}(X_i)}}. \]

---

4 Amitab Kundu et al, 2002 gives, Axiom R: Maintenance of Relativity, i.e. the scale transformation must not alter relative ranking of the observational units. Axiom S: Comparability and Standardization i.e. the mean of all the transformed indicators must be equal.

5 Gopal Krishan Pal, 1995; Savitri Abeyasekera; M Mukherjee and A K Roy, 1978; Noorbakhsh to name a few.

6 Pyatt et al., 1980, This gini coefficient is different from the population weighted gini coefficient.
where \( m = \left[ \sum_{j=1}^{k} \left( 1/\sqrt{\text{Var}(X_j)} \right) \right]^{-1} \)

\( X_j \) is the \( j^{th} \) variable, \( \text{Var}(X_j) \) is the variance in the variable, and \( k \) is the total number of variables.

The choice of the weights in this manner ensures that large variation in any one of the indicators will not unduly dominate the contribution of the rest of the indicators and distort inter district comparisons\(^7\).

Another method to assign weights to the variables includes the Principal Component Analysis (PCA). The present study employ PCA method to assign weights to the input variables for the construct of composite indices of development as the PCA method is considered both objective and non-arbitrary\(^8\). The weights in the PCA are determined by the product of the highest component loading of the variable from amongst the ‘p’ extracted components and the total variance explained by the respective component. In mathematical notation, it follows-

\[ w_j = a_j \sqrt{\lambda_p} \]

where, \( w_j \) represent the weight of the \( j^{th} \) variable, \( a_j \) represents the highest component loading of \( j^{th} \) variable on the \( p^{th} \) component that best explains the \( j^{th} \) variable. \( \sqrt{\lambda_p} \) is the total variation explained by \( k \)-variables in the \( p^{th} \) component. This procedure assigns more weights to variables accounting more variation.

In order to derive weights through PCA, we need to understand the basic concepts of components, component loadings, total variance explained by component/s in variable/s. To do so we need to understand factor analysis.

---

\(^7\) Refer Iyengar and Sudarshan, 1982 for taxonomic procedure and Noorbaksh (200...) for gini coefficient.

\(^8\) Refer Mukherjee and Roy, 1978.
A Brief on Factor Analysis: The Principal Component Approach (PCA)

Factor analysis is not a single statistical method, but represents a complex array of structure-analyzing procedures for a given concern, which may have several different but possibly interrelated sub-dimensions. Based on the interrelationships between the selected variables, it reduces the dimensionality of the original data set that summarizes and describes the original structural interrelationships in small number of ‘factors/components’ in the best possible way. A correlation matrix ‘R’ is used to summarize the interrelationships among variables in a given set. The determinant of correlation matrix ‘R’ ranges between 0 and 1 and it is critical to undertake further mathematical operations, either of the extreme cases is not desirable. On extremes, the correlation matrix becomes ‘ill-defined’ implying that some of the ‘eigen values’ associated with the correlation matrix are not positive or there may be as many factors as there are variables. A value close to zero indicates that there are several variables that are strongly correlated with one other (Marjorie A. Pett et al., 2003).

The Bartlett’s Test of Sphericity tests the correlation matrix. A chi-square test is used to test the null hypothesis that there is no relation among the variables. It takes the following form:

\[ \chi^2 = - \left[ (N-1) - \left( \frac{(2k + 5)}{6} \right) \right] \ln |R| \]

where \( \chi^2 \) is the calculated chi-square value for Bartlett’s test, N is the number of States, ‘k’ is the number of variables in the matrix, \( \ln \) is the natural logarithm and \(|R|\) is the determinant of the correlation matrix. The degree of freedom is,

\[ df = k (k-1) / 2. \]

If the calculated chi-square value is greater than the tabled chi-square value, we reject the null hypothesis of no interrelationships among the variables of the correlation matrix. Clearly, it is influenced by the sample size ‘N’. Large N gives large Bartlett’s values. If the Bartlett’s test is not found significant, the matrix is not fit to

carry out factor analysis. A more sophisticated test given by Kaiser, known as Kaiser-Meyer-Olkin (KMO) test, that compares the magnitude of the calculated correlation coefficients to the magnitude of the partial correlation coefficients and is calculated as given below:

$$
KMO = \frac{\sum (\text{correlation})^2}{\sum (\text{correlation})^2 + \sum (\text{partial correlation})^2}
$$

The KMO value range between 0 and 1, values close to zero are not preferred for factor analysis (FA). The criterion to select the variables for FA suggests that the KMO value should be at least 0.60 to derive meaningful results for FA/PCA. Also, since the correlation matrix of the variables that has zero determinant value is not considered for FA. The FA requires positively determinant of the correlation matrix for variables under consideration and thus requires sufficient number of correlations among variables. The numerator of the above mentioned statistic has to be greater than the denominator for positively defined correlation matrix. However, for the small sample size \(N<30\), Bartlett’s test provides a good indication for the quality of correlations among the variables. This is evident from the cautionary note suggested by Gorsuch, 1983, on the use of Bartlett’s test for large sample sizes (Marjorie A. Pett et al., 2003). With \(N=17\), less than desired, the present study largely depend on Bartlett’s test of sphericity to test the strength of relationships among the selected set of variables for factor analysis. Once we get the desired correlation matrix, we proceed further towards factor extraction procedure of factor analysis.

Among the different methods of extracting the factors, most commonly used are principal component analysis and principal factor analysis, while the other less used extraction methods include image factoring and maximum likelihood factoring, alpha factoring, unweighted least squares factoring, and generalized or weighted least squares factoring. The following discussion on what variances in the dataset are accounted for by the extracted factors will drive us to the use principal component analysis for this study.

It is found that the multivariate linear model of factor analysis break the sources of variation in ‘scores on a set of variables’. The broad divisions of the sources of variation are into the three uncorrelated components, namely ‘common variance’, ‘specific variance’, and ‘error variance’. Common variance explains the
shared variance among the variables and is presented by a set of common factors, commonly known as ‘communality’ and represented by \(h^2\). The specific variance of a variable is the variance in a variable not shared by other variables in a set. The unique or error variance represents the errors of measurement. The specific and error variance together form ‘unique variance’ symbolized by \(1-h^2\).

The factor extraction method that assumes extraction of total variance\(^ {10}\) explains the full component model and is known as ‘principal component analysis’. Thus the PCA is easy subject to criticism for its incapability of separating out the unique variance or for that matter the error variance. But it is widely used for its immense relevance in extracting uncorrelated components that extract variance in descending order and effectively summarizes the dataset.

We, in the present study would use PCA in R-mode\(^ {11}\) that linearly transforms the original set of \(k\) correlated variables into a substantially smaller set of \(p\) uncorrelated components, the principal components. These principal components are defined as linear combination of the original set of variables, such that the components retain maximum information and best explain the variance in the variables.

Let \(X_{ij}\)'s represents the original set of \(k\)-variables for \(N\) States, where \(i=1,2,\ldots,N\) and \(j=1,2,\ldots,k\), which in the standardized form is represented by \(Z_{ij}\)'s\(^ {12}\). The \(p\) uncorrelated principal components are represented by \(P_q\)'s, where \(q=1,2,\ldots,p\), and \(p<k\). The linear combination takes the following form:

\[
P_1 = a_{11} Z_1 + a_{12} Z_2 + \ldots + a_{1k} Z_k \\
P_2 = a_{21} Z_1 + a_{22} Z_2 + \ldots + a_{2k} Z_k \\
\vdots \\
P_p = a_{p1} Z_1 + a_{p2} Z_2 + \ldots + a_{pk} Z_k
\]

(B)

\(^{10}\) Total variance of a set of variables will give values equal to the number of variables, when variables are standardized for mean=0 and standard deviation=1.

\(^{11}\) In R-mode, the rows represent states and columns are variables, the cell entries are scores of the states on the variables and the derived components are the clusters of variables. The other modes of factor analysis are Q-mode, O-mode, T-mode, and S-mode. For more details visit http://www2.chass.ncsu.edu/garson/pa765/factor.htm.

\(^{12}\) \(Z_{ij}\)'s are derived form equation 2.

51
The above model represented by a set of equations ‘B’ is different from the full PCA model, which takes the following form:

\[

P_1 = a_{11} z_1 + a_{12} z_2 + \ldots + a_{1k} z_k \\
P_2 = a_{21} z_1 + a_{22} z_2 + \ldots + a_{2k} z_k \\
\vdots \\
P_p = a_{p1} z_1 + a_{p2} z_2 + \ldots + a_{pk} z_k
\]

\[(C)\]

In the full PCA model, represented by a set of equation ‘C’, the total number of extracted components is equal to the number of variables.

The a’s are the component/principal loadings such that the principal components P_q’s are orthogonal (uncorrelated) and the first principal component ‘P_1’ accounts for the maximum proportion of the total variance explained by the full PCA model B. The second component ‘P_2’ explains the maximum of the remaining variance and so on. The most sophisticated computer package, SPSS is used to compute principal loadings and the associated values of the PCA.

The principal loadings of ‘A’ are in fact the correlation coefficients that represent the magnitude as well as the direction of association between the variables and the principal components. Just as the squared correlation coefficients represent the proportion of variance common to the two variables, squared principal loadings represents the proportion of variance in a variable explained by a given component. For a given variable, the component with the highest (absolute) loading from amongst a set of extracted components, represented by a_j, thus best explains the variance in that variable. These ‘component loadings, a_j’ in association with the ‘total proportion variance explained, υ_p, by the respective components are used to determine weights to the variables, as presented in Equation 3 above. The total proportional variance of a given component, p, is derived in the following form:

\[
\gamma_p = \frac{\lambda_p}{k}
\]

\[(4)\]
where $\lambda_p$ is the eigen value of the $p^{th}$ component and $k$ is the total number of variables. The sum of the squared loadings for $p^{th}$ component is called its eigen value, $\lambda_p$.

Components with low eigen values contribute little to the explanation of variance in the variables and thus ignored. Thus, a set of ‘p’ components that best explain the interrelationships amongst a set of ‘k’ original variables is selected. A cut-off point to select the number of components to be included for the analysis is determined by eigen values. Eigen values equal to one or more determine the number of variables to be included in the study. The proportion of variance explained by the thus extracted set of ‘p’ components is given by:

$$\frac{\lambda_1 + \lambda_2 + \ldots + \lambda_p}{\lambda_1 + \lambda_2 + \ldots + \lambda_k},$$

where $p < k$, $\lambda$’s represents the eigen values of the principal components. For the full model of PCA, represented by set of equations ‘C’, $\lambda_1 + \lambda_2 + \ldots + \lambda_k = k$, where $k$ is the total number of variable involved.

The sum of the squared loadings of all the components, for a given variable explains the total variance in the variable that is accounted for by all the components. This is commonly known as ‘communality’ of a variable, and is denoted by $h^2$. It takes the following form:

$$h^2 = (a_{1j})^2 + (a_{2j})^2 + \ldots + (a_{kj})^2,$$

where $j = 1, 2, \ldots, k$

In the full model of PCA, set of equations ‘C’, $h^2 = 1$ (for $j = 1, 2, \ldots, k$), while in model B, $h^2 < 1$, (for $j = 1, 2, \ldots, p$ where, $p < k$). The communality extracted in model A explains only the extracted variance. Communality is thus the squared multiple correlation for the variable using the components as predictors and is interpreted as the measure of reliability of the indicator, and indicate how well each variable is explained by the extracted components. It is also important to note that the
squared loadings are added to get the proportion of variance explained only when the components are orthogonal.

Also, the error variance is taken care of when we restrict our analysis to include only the first few components that explain most of the variance, explained by model B. The critical value of 0.4 is considered as ‘low’ and misfits the variable to be a part of the model.

Though there are possibilities to determine weights through different approaches within the PCA model, the above explained procedure is followed. The other alternatives to determine weights to the variables may include:

1. All loadings on a variable of the ‘p’ extracted components may be considered to determine weights to the variables.
2. The squared loadings, the variance of a variable explained by each of the extracted components may also be considered to determine weights to the variables.

We shall use rotated components as the variables tend to weight more on one component in the unrotated PCA model which could make interpretation of components difficult. ‘Varimax criterion’ of orthogonal rotation would be used for the purpose.13

Given the weights to the variables, we need to construct a single aggregated measure that best represent the original variance in the dataset.

Methods for Constructing Composite Index from a given set of Standardized Variables

The weighted aggregated value for the \( i \)th State that captures best/maximum information from all variables in a given set, takes the following form:

\[
Y_i = \left( \sum_{j=1}^{k} (w_j z_{ij}) \right)
\]

13 Refer Rummel, www.hawaii.edu/powerkills
\[ Y_i = w_1 z_{i1} + w_2 z_{i2} + w_3 z_{i3} + \ldots + w_k z_{ik}, \quad (6a) \]
where \( Y_i \) is the weighted aggregated value of \( i^{th} \) State, \( w_j \) is the weight to the variables, and \( z_{ij} \) is the standardized value of \( j^{th} \) variable in the \( i^{th} \) State, and \( j=1,2,\ldots,k \).

And the composite index of development\(^{14} \) takes the following form-

\[ I_{at} = \frac{Y_{at}}{\sqrt[\sum_{j=1}^{k} (w_j)}} \quad (6b) \]

where \( \sum_{j=1}^{k} (w_j) = w_1 + w_2 + \ldots + w_k \) is the weight sum, \( Y_{at} \) is the weighted aggregated value at ‘t’ point of time, calculated from Equation 5, \( w_{jt} \) are the weights to the variables at ‘t’ point of time, the four points of time are 1971, 1981, 1991 and 2001. \( I_{at} \) is the composite index of the \( i^{th} \) State at the \( t^{th} \) point of time.

Thus the composite index of development would be calculated for each of the five pre-defined dimensions of development. The computed index values for each dimension are represented by \( I_{a1t}, I_{a2t}, I_{a3t}, I_{a4t}, I_{a5t} \) respectively. And, the composite index of overall development, computed from these five indices may be represented by, \( I_{ao} \).\(^{15} \) For the five dimensions, in respective sections of Chapter 4, these are introduced by name of the dimension it represents, say \( AD_{It} \) for Agriculture Development Index in \( t^{th} \) year. And for the composite index of development in Chapter 4, it read \( CID_{t} \).

\(^{14}\) For detailed understanding on various methods to construct composite index of development, refer the four main approaches discussed by Duncan, et al. 2004, the geometric mean index approach as applied by Hufter and Shah (1999), the arithmetic mean index approach as used by Manning, et al. (2000), and finally the principal component analysis approach as applied by Toatu (2002).

\(^{15}\) PCA seems to satisfy axioms for the relevance of the methods for giving weightage seems to be satisfied with PCA. Axiom C (a) An indicator having stronger interrelations with the other indicators should have higher weight, else (a’) Indicators that have weaker correlations must have higher weightage. Axiom D (b) An indicator having greater disparity in space must have a higher weight, else (b’) Indicators with greater dispersion in space should be given relatively smaller weights. Refer Kundu et al., 2002.
Since PCA is based on standardized z-scores that range between -3 to +3, the resultant PCA aggregate scores give both positive and negative values to the indices. The index values with positive sign represent above average States and those with negative sign represent below average States. The average of the index values across States remains zero. These original values of the index are given in the respective appendix.

For relative comparison of States we shall put all-India index value at 100 and examine distance of each State with base 100. These index values are represented by Index 1. This makes easy to compare States relative performance at a given point of time. However, to examine the trend in development of a State, we need to incorporate the changes at all-India level. The change at all-India levels of development over time would be introduced with all-India value of base year (1971) equal to 100. Thus, the indices in the following year will carry changes at all-India level. This new index is represented as Index 2. The inter State disparities would, thus, be examined for State indices on all-India 1971 base. The development trend of the States would be examined on the all-India 1971 base.

In order to examine the trend in the levels of development, we will use simple growth rate (in the development levels over a period of time divided by its base year value). The positive and negative sign to this growth rate would exhibit improvement or fall in development levels, respectively.

A construct called ‘catching-up’ would be deployed to examine how States fared over the years to cover their base year gap in development from the most competitive State in the base year, 1971. It takes account of development distance of each State from the most competitive State in the base year and examines how the percentage change in States own development levels over time helps to reduce its base year gap in development distance. The following algebraic notation is used to calculate the ‘catch-up’ rate for each State over a given period of time-

\[ Cr_i = \frac{[(I_{i,1} - I_{i,0}) \times 100]}{D_{0}}, \]

where \( Cr_i \) is the catching-up rate of the \( i^{th} \) State, \( I_{i,1} \) and \( I_{i,0} \) are the index values of the \( i^{th} \) State in

56
the current and base year respectively, \( D_o \) is the distance in the base year development level of the \( i^{th} \) State from the most competitive State of the base year.

**Slipage test** (M-statistics) will be used to examine the statistical significance of changes in the States development levels over time\(^{10}\). The equation, distributed as chi-square statistics with \( t-1 \) degree of freedom, estimates M-statistic for slippage test and takes the following form–

\[
M = \left[ 12/(N(t+1)) \right] \sum_{t=1}^{n} R_t^2 - 3N(t+1), \quad (8)
\]

where \( N \) is the number of States (\( N=17 \)), \( n \) represents four points of time \( t=1,2,...,n \) (\( n=4 \)), \( R_t \) denote sum of the ranks of the States at \( t^{th} \) point of time, and \( R_t^2 \) represent square of the sum of ranks in \( t^{th} \) point of time.

There might not be changes in relative positions of States over time but there may be changes in the extent of inter State disparities. The following measures shall be employed to estimate the extent of inter State disparities in the levels of development.

**Measures to Estimate Disparities/Inequalities**

Given the distribution, the average value provides crude information on the extent of development per State but does not tell us about how equal or unequal the values are across States. Thus, we shall depend to the following measures of dispersion.

Several statistical formulations relate to disparity, these include dispersion, skewness, and variance. The latter is also used as a measure of absolute convergence\(^{17}\). Different measures include range, range ratios, the Mc Loone Index, coefficient of variation (CV), and the gini coefficient. A brief discussion on each is given below.

\(^{10}\) Rai and Sarup, 1989.

\(^{17}\) For absolute convergence refer Chumacero, 2006.
Range (R)

Though simple to measure, Max ($X_i$) – Min ($X_j$) and easy to understand, its major limitation is that it gives all importance to the extreme values and ignore the underlying characteristics of the distribution, thus sensitive to outliers.

Range ratio (Rr)

Instead of using the extreme values as in Range, Rr determines a range of values on the top and a range at the bottom. It is thus the ratio of the two ranges measured as, average of the highest 10 percent (the 90th percentile) entries divided by the average of the bottom 10 percent (the 10th percentile). This is also known as the Decile Dispersion ratio. However, this ignores a great majority of the data.

Variance (Var)

This measure of dispersion is simply the average squared deviations, given by
\[ V = \frac{\sum (x_i - \mu)^2}{N} \]
where $\mu$ is the mean of the distribution, $x_i$ is the values of $i^{th}$ States for $j^{th}$ variable and $N$ is the total number of States.

It simply tells how far individual State is from the mean. The square values are used to avoid negative results.

Standard deviation (SD)

It is simply the square root of variance, $SD = \sqrt{Var}$, \hspace{1cm} (9)

This brings the square values back to normal. However, both the variance and the standard deviation measures depend on units of measurement.

The coefficient of variance (CV)

This measure is the ratio of standard deviation and the mean, represented as-
\[ CV = \frac{SD}{\mu} \]
\hspace{1cm} (10)

The ratio makes it unit free. This is its distinct quality for which it has accumulated sound support for its application on any set of variables for comparable analysis. Ceteris paribus, the smaller the coefficient of variation, the more equitable is the distribution. We use this measure for our analysis on inter States disparities.
It is based on full information in the distribution, and can take any value between zero and infinity. There is no universal standard that defines a reasonable value of the measure for particular phenomena\(^{18}\).

**The Gini coefficient (g)**

Its simple algebraic formula requires the data set (States) to be first ordered from smallest to largest values of the variable. It takes the following form:

\[
g_{j} = \frac{\left(\sum (2r_{i} - N - 1)X_{ij}\right)}{(N^{2} \mu)}, \tag{11}
\]

where \(g_{j}\) is the Gini coefficient of the \(j^{th}\) variable, \(r_{i}\) is the ranks of the \(i^{th}\) State, \(N\) is the total number of States, \(X_{ij}\) is the values of the \(j^{th}\) variable on the \(i^{th}\) State, and \(\mu\) is the mean of the distribution.

The coefficient varies between zero (which reflects complete equality) and 1 (which indicates complete inequality) and can be easily represented by the area between the Lorenz curve and the line of equality. The only disadvantage of the Gini coefficient is that it is non-additive across groups. The other important measures of inequality are Theil Index and Atkinsons Index, but this study depends on CV and Gini coefficient.

The neoclassical economic theory predicts the converging of regional income if all regions share similar steady states\(^{19}\). We test our results on development of States for convergence over 30 year time period, and examine how long the State will take to reach half-life to steady state. The convergence analysis is also made for the two sub-period representing pre and post reform periods, separately. The study largely employs the three measures of dispersion, coefficient of variance, Gini coefficient, and standard deviation in logarithm index values which is considered as a measure of absolute convergence.

**Measuring Convergence**

The two standard measures of ‘unconditional/absolute convergence’ known as ‘\(\sigma\) (sigma) convergence’ and ‘\(\beta\) (beta) convergence’ are tested for PC NSDP at constant 1993-94 prices and development levels across States over the period 1971-

\(^{18}\) Refer Hale, 2004.

\(^{19}\) Barro, 1991.
The test for conditional convergence assumes control on factors responsible for determining long run growth rates.

**Absolute $\sigma$-Convergence**

The standard deviation (SD) of the natural logarithm of PC NSDP ($\ln$ PC NSDP) is used for testing ‘$\sigma$–convergence’. There is $\sigma$–convergence if the standard deviation tends to decline over time. It speaks straightforwardly to whether the distribution of income across States over time is becoming more equitable or not.

**Absolute $\beta$–Convergence**

The ‘$\beta$–convergence’ is a primary focus of growth empirics because it serves a necessary condition for ‘$\sigma$–convergence’\(^{20}\). The possibility for convergence is determined with ‘$\beta$–convergence’ though ‘$\beta$–convergence’ may not bring ‘$\sigma$–convergence’. The following regression framework in a simple univariate regression equation form\(^{21}\) is used to examine ‘$\beta$–convergence’.

\[
\Delta Y = \alpha + \beta Y_0 + u, \quad (12)
\]

where, $\Delta Y$ is the change in $\ln$ values of $X$ ($\ln X_t - \ln X_0$), $Y_0$ is the $\ln$ value of $X$ in the base year ($\ln X_0$), and $u$ is the error term that assumes to satisfy the Gauss Markov assumption.

Thus we regress the logarithm of change in ‘$X$’ over time on its initial value. Negative and statistically significant $\beta$-coefficient suggests $\beta$–convergence.

**Conditional $\beta$–Convergence**

The conditional convergence hypothesis is tested when States possess the same initial conditions, in income and other factors influences long run growth rate, there is still a possibility for convergence to the same growth rate which may not be accomplished at the same steady state growth rates. Following numerous studies testing conditional convergence\(^{22}\), an additional explanatory variable, $X$ is introduced

\(^{21}\) Chumacero, 2006.
\(^{22}\) Chumacero (2006); Young, et al. (2004); Barro et al (1992); and others.
to the univariate regression Equation 12 to devise regression for conditional convergence. The resultant equation takes the form—

\[ \Delta Y = \alpha + \beta Y_0 + \theta X + u, \]  

(13)

where, \( \Delta Y \) is the change in \( \ln \) values of \( X \) (\( \ln X_t - \ln X_0 \)), \( Y_0 \) is the \( \ln \) value of \( X \) in the base year (\( \ln X_0 \)), \( X \) is the \( \ln \) value of the conditional variable in the base year.

The positive or negative sign of the estimated ‘\( \beta \)’ and ‘\( \theta \)’ coefficients establish the link that speaks for divergence or convergence respectively. It is important to note that the absolute convergence is favored by the data if the estimated ‘\( \theta \)’ is negative and statistically significant (different from zero). The ‘\( \beta \)’ coefficient explains the rate of convergence.

**Speed of Convergence and Half life to the Balanced Growth**

We also attempt to measure the speed of convergence. It explains by what annual growth rate economies will grow to close the gap between their present level and their balanced growth/steady state. It is measured as,

\[ \lambda = -[\ln (1+b)] / T. \]  

(14)

where \( \lambda \) is the speed of convergence, ‘\( b \)’ is the estimated \( \beta \) coefficient (the rate of convergence), and \( T \) is the number of years over which estimation is to be made.

The half life, is the time necessary to cover half distance to the steady state and is measured as,

\[ t = - \ln (0.5) / \lambda, \]  

(15)

where \( t \) is the half life to reach steady state, and \( \lambda \) is the speed of convergence.

---

The study tests convergence in per capita net state domestic product as well as in development levels of the selected 17 major Indian States. Since the States within nations have less or negligible barriers to the movement of goods and services as compared to international barriers, the spatial spillovers are expected to be more and may result in dependence of weak States on stronger neighboring States. Testing State level data for spatial dependence and convergence analysis for autocorrelation may provide more accurate results. The spatial dependence of States in India presumes that the States growing fast are those near or surrounded by the already developed States.

The per capita net state domestic product (PC NSDP) is tested for spatial dependence of States and the spatial dependence in PC NSDP, if any, is introduced in the convergence analysis with spatial lag regression model. The differences in the convergence results with and without spatial lag determine the significance of spatial dependence in income across States, and thus help to decide if the spatial lag is also to be introduced in examining convergence for development levels across States.

**Testing Spatial Auto Correlation through Moran I Statistic**

The local Moran test (the Moran's I statistic) is used to detect local spatial autocorrelation, that is, to identify local clusters (regions where adjacent areas have similar values) or spatial outliers (areas distinct from their neighbors). Moran's I statistic, M(I), a conventional measure of autocorrelation is calculated at each point of time to test the hypothesis that there is no association between the value observed at a location and values observed at nearby sites, that is, values of M(I) is close to zero. It is measured as-

\[
I_i = \frac{n}{S} \cdot \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} z_i z_j \right\} \left/ \frac{1}{n} \sum_{i=1}^{n} z_i^2 \right\}
\]

(16)

where 'n' is the number of States, 'S' is the sum of the elements in the spatial weight matrix W which summarizes the spatial effects between States. \(w_{ij}\) are the elements of the spatial weight matrix.
matrix $W$ corresponding to the regions $i$ and $j$, and $z_i$ and $z_j$ are normalized vectors of the log of per capita NSDP of States $i$ and $j$ respectively.

Moran’s $I$ statistic can take values between $-1$ and $+1$. Values close to $+1$ represent strong and positive spatial dependence, clustering of similar values, whereas values around $-1$ show negative spatial correlation, clustering of different values. The $z_i = \ln \left( \frac{\text{PC NSDP}_i}{\text{PC NSDP}} \right)$ denotes the logarithm of the PC NSDP of State $i$ in period $t$, (PC NSDP$_i$) normalized by the total of all States sample mean of the same variable, PC NSDP.

The contiguity spatial weight matrix `$W$‘ is a binary matrix of $n \times n$ order (17 x 17 in our case of 17 States). It detects spatial dependence, it takes the value of 1 if two States $i$ and $j$ are neighbors and 0 for other entries of the matrix, and for $w_{ij} = w_{ji} = 0$.

If the PC NSDP values are distributed normally, then $I$ is also assumed to be normally distributed with expected value, $E(I)$ and variance, $\text{var}(I)$ given as:

$$E(I) = \frac{1}{(n-1)}$$

$$\text{Var}(I) = \frac{n\{(n^2 - 3n + 3)S_1 - nS_2 + 3S_0^2\} - k\{n(n-1)S_1 - 2nS_2 + 6S_0^2\} - 1}{(n-1)^2}$$

Where,

$$S_1 = \frac{1}{2}\left\{ \sum_i \sum_{j \neq i} (w_{ij} + w_{ji}) \right\}$$

$$S_2 = \sum_i (w_{ij} + w_{ji})^2$$

and $w_{ij} = \sum_j w_{ij}$, $k = m_0^i / m_0^j$ with $m_0 = 1/ n \sum_i (x_i - \mu)^2$.

$$Z = \frac{[M(I) - E(I)] / \text{SD}(I)}{\text{SD}(I)} = \text{Square root of Var}(I).$$

Moran’s scatterplots may provide a closer look at the pattern of spatial concentration. The plot features standardized values for each State against its spatial lag value; the standardized values are on x-axis and the mean standardized neighbor values on y-axis, since the values are in standardized form the values correspond to the

---

27 Appendix A 3, Table A 3.1 for 17x17 binary contiguity spatial weight matrix ‘$W$’ for States under consideration in this study.

standard deviations. The scatterplot is divided into four quadrants providing classification of four types of spatial autocorrelation.

Areas that are significant are labelled with ‘high-high’ (quadrant I) or ‘low-low’ (quadrant III) categories produced in the Moran scatterplot analysis. High clustering of States in these two quadrants provide for significant spatial autocorrelation.

Since the traditional approach to empirically test convergence is based on cross-section regressions, the residuals must satisfy standard Gauss-Markov assumptions to have meaningful cross-section regression results. If there is spatial autocorrelation in the State level data, then the residuals of the regression may also be high on spatially autocorrelation, which violates the Gauss Markov assumptions. In that case, the estimate of convergence parameter ‘beta’ may not be reliable.

---

29 For any linear regression to have Best Linear Unbiased Estimates, the Gauss-Markov assumptions states that the expected value of the error term should be ‘zero’, error terms have same variances and are uncorrelated that is \( \text{cov}(e_i,e_j)=0 \).

30 Badinger et al., 2002
Therefore, existence of spatial autocorrelation in the residuals of the regression also needs to be examined before we proceed further.

A number of test statistics are suggested in the literature to test spatial autocorrelation in the residuals. For a row standardized spatial weight matrix, the most commonly used test statistic for spatial autocorrelation in regression residuals, the Moran’s I statistic takes the following form:

\[ I = \frac{(e'W)e}{(e'e)} \]  \hspace{1cm} (17)

where ‘e’ is the vector of residuals from OLS regression, e’ is its transpose, W is the contiguity spatial weight matrix.

In the presence of spatial autocorrelation, the easiest model used to integrate it in to the convergence analysis is the spatial cross-regressive model. A variable for spatial lag is introduced in the original convergence equation to account for spatial dependence across States. The model thus is written as:

\[ \Delta Y = \alpha + \beta Y_0 + \delta WY' + u, \]  \hspace{1cm} (18)

where, \( \Delta Y \) is the vector of change in ln values of \( X \) (\( \ln X_t - \ln X_0 \)), \( Y_0 \) is the vector of ln value of \( X \) in the base year (\( \ln X_0 \)). \( WY' \) is the spatial lag vector which averages the \( \ln X_0 \) of the neighboring States, and \( u \) is the error term that satisfies the Gauss-Markov assumptions.

Since the spatial lag of the independent variable is exogenous, the model can be estimated via OLS. Comparing the convergence results of spatial lag model with the conventional model of absolute beta convergence, we determine the impact of spatial dependence on our original conversion results.

---

31. Row standardized spatial weight matrix imply that the weight matrix is such that sum of each row of the matrix equals 1. Moran’s I, the standard test for spatial autocorrelation in regression residuals is calculated as: \[ I = \frac{n}{S}, \\frac{(e'W)e}{e'e}, \] where e is a vector of residuals, n is the number of observations, and S is a standardization factor equal to the sum of all the elements in the weight matrix. For a row standardized spatial matrix \( n=S=17 \), therefore \( n/S=1 \).
33. Badinger et al., 2002.
The study also attempts to determine a smaller set of variables that best summarize development across States. This is done through stepwise regression for individual and composite indices of development. A smaller set of variables so determined are important to suggest specific strategies to the underdeveloped States to help them catch the race in the process of development. The provisional time estimates for the States development to reach steady state growth rates may help them to formulate their future growth strategy.

**Stepwise Regression Analysis**

Stepwise regression first finds the most correlated explanatory variable. A variable with highest partial correlation with the dependent variable is then added controlling for the previous. The process goes on till additional variable significantly increase R-square ($R^2$).

The $R^2$ values determine the proportional variation in the dependent variable explained by a given set of variables. It is calculated as the ratio of explained (regression) sum of squares to the total sum of squares. The statistical significance of the relationship between this independent variable and the dependent variable (controlling for the other variables in the model) is calculated by t-values, where $t$ is the ratio of regression coefficient and its standard error.

The associated p-values to the given t-values determine the level of significance. The small p-values provide significant contribution of explanatory variable, significant F change, and significant increase in $R^2$. The p-values less than 0.01 give 99 per cent significance level, less than 0.05 give 95 per cent significance level, and less than 0.10 give 90 per cent significance level.

The statistical significance of the model, with k number of explanatory variables for N States is determined by F-value, which takes the form–

$$F = \frac{[R^2 / (1 - R^2)] [(N - k - 1)/k]}{N - k - 1}$$

with $N - k - 1$ degree of freedom.

The significant F-change provides significant change in $R^2$. The Adjusted $R^2$ takes into account not only how much of the variation is explained, but also the impact
of the degree of freedom. It ‘adjusts’ for the number of variables in use, to see how adding another variable to the model both increases the explained variance but also lowers the degrees of freedom.

\[
\text{Adjusted } R^2 = 1 - (1-R^2) \times \left( \frac{(N-1)(N-k-1)}{N-1} \right)
\]

As the number of variables in the model increases, the gap between the \(R^2\) and the adjusted \(R^2\) increase. The standardized regression coefficients in the regression analysis determine the relative importance of variables.

The above discussed statistical procedures that serve wide ranging needs of the study are used in the following chapters.

**SOURCES OF DATA**

The authenticated sources of secondary data/information provide strong base to this study. The list includes:

- **Census of India**
  - (i) India- A Statistical Outline, various issues;
  - (ii) Tables on Houses, Household Amenities and Assets, various issues;
  - (iii) Various Issues of Primary Statistical Abstracts, and Socio Cultural Tables;

- **Central Statistical Organization**
  - (i) Economic Survey, various issues;
  - (ii) Statistical Abstract, India, various issues;
  - (iii) Statistical Abstracts, Punjab, various issues;
  - (iv) Statistical Abstracts, Haryana, various issues.

- **Economic and Political Weekly Research Foundation**

- **National Council on Education and Applied Research**
  - (i) All-India Education Survey, various issues.
Reserve Bank of India
(i) Report on Trend and Progress of Banking in India, various issues;
(ii) Banking Statistics, Basic Statistical Returns, Summary Results, various issues.

Sample Registrar System

The CIA Fact book
https://www.cia.gov/cia/publications/factbook

The Fertiliser Association of India
(i) Fertiliser Statistics, various issues, FAI, New Delhi.

The Planning Commission
(i) The Five Year Plan Documents, various issues, Planning Commission, New Delhi;

The World Bank
(i) World Development Indicators, various issues

The selected statistical and econometric applications are applied in the following Chapters.