3.1 INTRODUCTION

The present chapter empirically measures the pattern and scope of rural transformation in India through a broad evaluation system based on the different statistical techniques. Previous investigation on rural transformation showed that rural India has witnessed a great transformation in period of strong economic growth. This transformation of India economy has also improved the condition of rural areas. But, some large regional disparities in the transformation of rural areas are still evident. The earlier investigators reported that strategies aimed at specific transformation types of the region in rural areas could be an useful way to shape a more integrated pattern of urban-rural development in India.

Rural transformation is a more dynamic concept than rural development, since it represents a change in the lifestyles of people (Shaw, 2011). It is generally characterized by changes in employment structure, agricultural income, agricultural intensity, public facilities, gender ratio, female literacy, crop selection, labor productivity and significant changes in rural housing and socio-economic conditions due to industrialization and urbanization (Curien, 1980; Ravallion and Datt, 2002). Various econometric methods had been used to analyze the rural transformation by different researchers such as co-integration, principle component analysis, ginni index, correlation analysis and regression analyses. Each method has its own significance to get the results.

In the present study, different statistical analysis has been conducted to fulfill the objectives. This analysis can provide relevant statistical information for better understanding of indicators of rural transformation and effect of urbanization on rural transformation in India as well as Haryana state.

As data are the basic and preliminary requirement for conducting any type of research study and the reliability and accuracy of the results also depends upon it. A brief description about the data and sampling design has been presented in this section. This section also deals with the analytical tools adopted to achieve objectives of the study.

3.2 SAMPLING DESIGN

Two stages purposive sampling technique has been used to select the districts and villages for the study.

3.2.1 Selection of Districts

Haryana consists of 21 districts and 6841 villages (Census, 2011). Out of these 21 districts, three districts namely Mahendergarh, Rewari and Gurgaon of Haryana has been selected purposively due to their strategic location to assess the effect of urbanization on employment, wages and livelihood in these districts. These districts have been selected because the effect of urbanization could be reflected more accurately with these districts as Mahendergarh is second least urbanized while Gurgaon is second most urbanized district (Census 2011). On the other hand Rewari comes under middle urbanized districts of state. Thus, these districts are the most suitable districts for studying the effect of urbanization on rural transformation in the Haryana state.

3.2.2 Selection of Villages

The selected districts consist of 1015 villages. To diagnose the effect of urbanization on employment, wage and rural livelihood, all on road villages near to urban centre (up to 10 Km) and away from urban centre (more than 10 Km) have been selected randomly with assumptions that villages near to urban centre may have benefitted more by urbanization as compared to villages located away from urban centre. The head quarter of each selected district has been considered as the urban centre. The villages located up to 10 Km from district headquarter on Mahendergarh - Gurgaon road has been considered as near to urban centre and villages located more than 10 Km from district head quarter have been considered away from urban centre irrespective of the district in which village is located. Thus, villages located either side of national highway/state highway/district roads from Mahendergarh – Gurgaon road have been selected for study. The total of 90 villages located either side of national highway/state highway/district roads from Mahendergarh – Gurgaon road have been selected. The village head and panchayat members of the selected villages have been approached to collect required information.

3.3 DATA COLLECTION

The present study is based on both primary and secondary data. The primary and secondary data have been collected from primary and secondary sources.

3.3.1 Primary Data

To assess the effects of urbanization on employment, wages and livelihood of rural people in Haryana, primary data has been collected.

The primary data has been collected during the month of June and July, 2017 from the sampled villages by conventional survey method using well structured schedule through personal interview. The data on distance of village from urban centre, population of village, number of households in village, demographic characteristics of population in village, workers, industry of their employment, price of commodities, wages for various works, migration pattern of people, reasons for migration, perception of village head/panchayat members about effect of urbanization on the proximity to urban area, better market facility, better services, better job opportunities, income earning opportunity, availability of NGO/SHG in the village, lack of facilities, role of agency in village development, etc, has been collected to study the effect of urbanization on employment, wages and rural livelihood.

3.3.2 Secondary Data

To ascertain the determinants of rural transformation in India, secondary data has been used. There are many indicators of rural transformation, but after reading national and international literature, it has been found that rural unorganized manufacturing industries, urbanization, rural road, rural literacy, GDP of agricultural & allied sector are the important indicators of rural transformation. Therefore, the secondary data on rural unorganized manufacturing industries, urbanization, rural road, rural literacy, GDP of agricultural & allied sector for 23 states (20 states and 3 Union Territories) have been considered from various published/unpublished sources. In case of Haryana, the data on above variables was available only for fifteen districts. The sources and period of secondary data has been presented below in table 3.1.

Table 3.1

Data, Period of Data and Their Sources

Sl. No.	Description of data	Period of data	Source
1	Unorganized manufacturing industries (for state & districts)	1994-95, 2000-01, 2005-06 and 2010-11	NSSO reports & unit level data on unorganized manufacturing
2	Rural non-farm employment (for state)	1993-94, 1999-00, 2004-05 and 2009-10	NSSO reports & unit level data employment & unemployment
3	Rural non-farm employment (for districts)	1991, 2001 and 2011	Population census of India of relevant periods
4	Rural road in Km. (for state)	1994-95, 2000-01, 2005-06 and 2010-11	Basic road statistics of relevant periods
5	Urban population (for state & districts)	1991, 2001 and 2011	Population census of India of relevant periods
6	Rural literacy (for state)	1991, 2001 and 2011	
7	Agricultural & allied gross domestic products (for state)	1994-95, 2000-01, 2005-06 and 2010-11	DES, Ministry of Agriculture & Farmers' Welfare, Government of India

As, it is evident from table, the data on relevant variables was not available for uniform period. Thus, the data on relevant variable has been predicted for the period 1994-95, 2000-01, 2005-06 and 2010-11 using methodology suggested by the experts of Institute of Economic Growth (IEG), Delhi to make the period uniform for further analysis.

Since, the data collected from various secondary sources on relevant variables was in absolute term and using these variables in such form may result in biasness towards bigger state because they may hold larger proportion of all or some variables. To overcome such problem, variables have been transformed in the manner shown in table 3.2.

Table 3.2: Description of Variables

Sl. No.	Original variable	Transformation of Variables
1	Rural non-farm employment (RNFE)	$\frac{\text{Total RNFE}}{\text{Total rural working population}}$
2	Urbanization (Utp)	$\frac{\text{Total urban population}}{\text{Total population}}$
3	Rural unorganized manufacturing industries (Um)	$\frac{\text{Total rural unorganized manufacturing industries}}{\text{Total unorganized manufacturing industries}}$
4	Rural road in km. (Rrp & Rrg)	$\frac{\text{Total rural road}}{\text{Total rural population}}$ $\frac{\text{Total rural road}}{\text{Geographical area of state}}$
5	Agricultural & allied sector GDP (Agnsa & Agrp)	$\frac{\text{Agricultural \& } allied\ sector\ GDP}{Net\ sown\ area}$ $\frac{\text{Agricultural \& } allied\ sector\ GDP}{Total\ rural\ population}$
6	Literacy (Lr)	$\frac{\text{Total rural literate population}}{\text{Total rural population}}$

3.4 ANALYTICAL TOOLS

The analytical tools and techniques used to derive results of each objective have been described in this section. The methodology found to be most suitable on the basis of review of literature and suitability with respect to data and time has been used.

3.4.1 Determinants of Rural Transformation

To study the determinants of rural transformation in India, panel data regression model has been employed. The transformed data (as shown in table 3.2) has been used for panel data regression. The panel data regression model with different combinations of explanatory variables has been used to determine the most suitable regression model, to ascertain the determinants of rural transformation. The general form of panel data regression analysis is given below:

$$\boldsymbol{Y_{it} = a_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \beta_4 X_{4it} + \beta_5 X_{5it} + u_{it}}$$

Where,

Y_{it} = Rural non-farm employment in i^{th} state in t^{th} time period.

X_{1it} = Unorganized manufacturing industries in i^{th} state in t^{th} time period.

X_{2it} = Urban population in i^{th} state in t^{th} time period.

X_{3it} = Length of rural road in i^{th} state in t^{th} time period.

X_{4it} = Rural literacy in i^{th} state in t^{th} time period.

X_{5it} = Agricultural & allied sector GDP of i^{th} state in t^{th} time period on 2004-05 base year.

The panel data regression model with different combination of independent/explanatory variables used have been presented below:

Model 1 $RNFE_{it} = \beta_0 + \beta_1 Um_{1it} + \beta_2 Utp_{2it} + \beta_3 Lr_{3it} + \beta_4 Rrg_{4it} + \beta_6 Agrp_{5it} + u_{it}$

Mode 2 $RNFE_{it} = \beta_0 + \beta_1 Um_{1it} + \beta_2 Utp_{2it} + \beta_3 Lr_{3it} + \beta_4 Rrg_{4it} + \beta_7 Agnsa_{5it} + u_{it}$

Model 3 $RNFE_{it} = \beta_0 + \beta_1 Um_{1it} + \beta_2 Utp_{2it} + \beta_3 Lr_{3it} + \beta_5 Rrp_{4it} + \beta_7 Agnsa_{5it} + u_{it}$

Model 4 $RNFE_{it} = \beta_0 + \beta_1 Um_{1it} + \beta_2 Utp_{2it} + \beta_3 Lr_{3it} + \beta_5 Rrp_{5it} + \beta_6 Agrp_{6it} + u_{it}$

Model 5 $RNFE_{it} = \beta_0 + \beta_1 Um_{1it} + \beta_2 Utp_{2it} + \beta_3 Lr_{3it} + \beta_4 Rrg_{4it} + \beta_5 Rrp_{5it} + \beta_6 Agrp_{6it} + \beta_7 Agnsa_{7it} + u_{it}$

These five models have been used to ascertain the robustness of the results to be obtained by panel data regression analysis.

Apart from the above model, panel data regression model with fixed and random effect have also been estimated to ascertain the factors responsible for rural transformation. The Hausman test had been used to ascertain the most suitable panel data regression model between fixed and random effects model.

3.4.1.1 Fixed Effect Approach

The basic assumption of fixed effect model is that although the intercept (β_{0i}) may differ across individuals but each individual's intercept does not vary over time, i.e., it is time invariant. It also assumes that the slope coefficients (β_{it}) of the explanatory variables do not vary across individuals or over time. Panel data regression model with fixed effect of following nature has been estimated.

$$\mathbf{RNFE_{it}} = \boldsymbol{\beta_{0i}} + \boldsymbol{\beta_1}\mathbf{Um_{1it}} + \boldsymbol{\beta_2}\mathbf{Utp_{2it}} + \boldsymbol{\beta_3}\mathbf{Lr_{3it}} + \boldsymbol{\beta_4}\mathbf{Rrg_{4it}} + \boldsymbol{\beta_5}\mathbf{Rrp_{4it}} + \boldsymbol{\beta_6}\mathbf{Agrp_{5it}} + \boldsymbol{\beta_7}\mathbf{Agnsa_{6it}} + \boldsymbol{\varepsilon_{it}}$$

3.4.1.2 Random Effect Approach

The panel data regression model with random effect of following nature is also estimated to ascertain the determinants of rural transformation:

$$\mathbf{RNFE_{it}} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}\mathbf{Um_{1it}} + \boldsymbol{\beta_2}\mathbf{Utp_{2it}} + \boldsymbol{\beta_3}\mathbf{Lr_{3it}} + \boldsymbol{\beta_4}\mathbf{Rrg_{4it}} + \boldsymbol{\beta_5}\mathbf{Rrp_{5it}} + \boldsymbol{\beta_6}\mathbf{Agrp_{6it}} + \boldsymbol{\beta_7}\mathbf{Agnsa_{7it}} + \boldsymbol{\varepsilon_{it}}$$

Where,

RNFE = Ration of rural non-farm employed person to total working population of the state,

Um = Ratio of rural unorganized manufacturing industries to total unorganized manufacturing industries in the state,

Utp = Ratio of urban total population to total population of the state,

Lr = Ratio of rural literate persons to total rural population of the state,

Rrg = Rural road per Km. of geographical area of state,

Rrp = Rural road per rural person in the state,

Agrp = Agricultural & allied GDP per rural person in the state, and

Agnsa = Agricultural & allied GDP per hectare net sown area in the state.

3.4.1.3 Hausman's Test

The specification test devised by Hausman (1978) "for the orthogonality of the random effects and the regressors is based on the concept that under no correlation hypothesis, both OLS in the fixed effect model and GLS are consistent, but OLS is inefficient, whereas under the alternative, OLS is consistent, but GLS is not. Thus, under the null hypothesis, the two estimates should not differ systematically and a test can be based on the difference." The other essential ingredient for the test is the covariance matrix of the difference vector, $[\mathbf{b} - \hat{\boldsymbol{\beta}}]$:

$$\mathbf{Var}\ [\mathbf{b} - \hat{\boldsymbol{\beta}}] = \mathbf{Var}\ [\mathbf{b}] + \mathbf{Var}\ [\hat{\boldsymbol{\beta}}] - \mathbf{Cov}\ [\mathbf{b}, \hat{\boldsymbol{\beta}}] - \mathbf{Cov}\ [\mathbf{b}, \hat{\boldsymbol{\beta}}]$$

Hausman's essential result is that the covariance of an efficient estimator with its difference from an inefficient estimator is zero, which implies that

$$\mathbf{Cov}\ [(\mathbf{b}, \hat{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}}] = \mathbf{Cov}\ [\mathbf{b}, \hat{\boldsymbol{\beta}}] - \mathbf{Var}\ [\hat{\boldsymbol{\beta}}] = \mathbf{0}$$

or that

$$\mathbf{Cov}\ [\mathbf{b}, \hat{\boldsymbol{\beta}}] = \mathbf{Var}\ [\hat{\boldsymbol{\beta}}]$$

This result produces the required covariance matrix for the test,

$$\mathbf{Var}\ [\mathbf{b} - \hat{\boldsymbol{\beta}}] = \mathbf{Cov}\ [\mathbf{b}, \hat{\boldsymbol{\beta}}] - \mathbf{Var}\ [\hat{\boldsymbol{\beta}}] = \boldsymbol{\psi}$$

The chi-squared test is based on the Wald criterion:

$$\mathbf{W} = \mathbf{X}^{2}\ [\mathbf{K\text{-}1}] = [\mathbf{b} - \hat{\boldsymbol{\beta}}]\ \hat{\boldsymbol{\psi}}^{-1}\ [\mathbf{b} - \hat{\boldsymbol{\beta}}]$$

For $\hat{\boldsymbol{\psi}}$, the estimated covariance matrices of the slope estimator in the fixed effect model and the estimated covariance matrix in the random effects model, excluding the constant term was used. Under the null hypothesis, W has a limiting chi-squared distribution with K−1 degrees of freedom.

The Hausman's test for the fixed and random effects regressions is based on the parts of the coefficient vectors and the asymptotic covariance matrices that correspond to the slopes in the models, that is, ignoring the constant term(s).

3.4.2 Association between Urbanization, Rural Industrialization, Industrialization and Rural Livelihood in Haryana

The association of urbanization, industrialization, rural-industrialization and livelihood in Haryana has been studied using Pearson correlation coefficients (r). In study, rural non farm sector (RNFE) was considered as the indicator of rural livelihood as it increases both the income and employment. Therefore secondary data has been collected for 15 districts of Haryana to check this association The Pearson correlation coefficient measures the strength of association between two variables. The formula of Pearson correlation coefficients is given below:

$$\boldsymbol{r} = \frac{\boldsymbol{n}(\sum \boldsymbol{xy}) - \sum \boldsymbol{x} \sum \boldsymbol{y}}{\sqrt{[\boldsymbol{n}\sum \boldsymbol{x}^2 - (\sum \boldsymbol{x})^2]\,[\boldsymbol{n}\sum \boldsymbol{y}^2 - (\sum \boldsymbol{y})^2]}}$$

Where,

r = Pearson correlation coefficient

x = First variable, and

y = Second variable.

n = Number of observations

After arriving at correlation coefficient, t-test is used to test the statistical significance of strength of association between variables under consideration.

Before applying the above analytical tools (panel data regression model and Pearson correlation coefficient), the data collected from various different sources (state level and district level) has been predicted for required period to make the period uniform for all variables, because data available on relevant variables were of different period. The prediction of data has been carried out in three steps after discussion with the expert available in the Institute of Economic Growth (IEG), Delhi. The formula used for predicting the data has been given below.

Step 1: First of all annual growth rate was computed in following manner

$$\textbf{Annual growth rate (R)} = \left[\frac{\text{Log }(V_{t+1})/\log(V_t)}{N}\right]$$

Step 2: Thereafter, value was estimated

$$\textbf{Estimated value} = \log(V_{t+1}) + (N-0.5) \times R$$

Step 3: The antilog of estimated value has been taken to arrive final projected value for respective variables, i. e.,

Antilog (estimated value) = Projected value

Where,

V_t = Observed value of variable under consideration in t^{th} period

V_{t+1} = Observed value of variable under consideration in t+1 period

N = Gap in the period for which data was not available

3.4.3 Effect of Urbanization on Employment, Wages and Rural Livelihood

The tabular analysis was used to study the effect of urbanization on employment, wages and rural livelihood. The simple statistical tools as percentage, average, etc. have been used to arrive at the results of this objective.

3.4.4 Reliability Test

Reliability is considered to be an important element of measuring the quality of test. Reliability is defined as how well a test measures what is should. 'Cronbac's Alpha Test' has been used to assess the reliability of the schedule. Cronbac's Alpha was developed by Lee Cronbach in 1951. It measures the reliability or internal consistency. It is denoted by **α** (alpha) and when inter-correlation among tested items increases, the value of **α** will also increase. Theoretically, it varies from 0 to 1. Many researchers recommend that accepted value of **α** lies between 0.65 to 0.8 (or higher in many cases), **α** values that are less than 0.5 are generally undesirable.

SPSS software was used to find the value of reliability statistics under Cronbac Alpha test for the primary survey. The value of **α** was 0.70, shows a good level of reliability. The result of Cronbac's Alpha test is given below:

Reliability Statistics	
Cronbach's Alpha	No. of Items
.707	72