STATISTICAL ANALYSIS ON MIGRANTS USING MARKOV CHAIN MODEL

[1]G. Venkatesan and [2]V. Sasikala
1. Associate Professor and Head, Department of Statistics,
2. Ph.D., Research Scholar, Department of Statistics,
Government Arts College (Autonomous), Salem, Tamilnadu, India

Abstract: India is an agricultural country and it has the second largest population in the world. In 2011, it is about 17.64% of the world population. The ratio of urban population is less than rural population. Ratio of urban populations to the total population of a country is an index of the level of industrialization of that country. Population policy refers to improve the quality of life. The main cause of rise in urban population in India is "Migration Effect". We estimate the Rural to Urban migration using markov chain analysis. This paper attempts various theoretical concepts of demography and migration. On using census data with relevant information from various sources we estimate the rural urban migration distribution using markov chain model for India. We concluded that migrant effect is also an index of urbanization.

Key words: Demographic process, Migration, Census data, Markov chain.

1. Introduction:

Demographic process includes Fertility, Mortality and Migration. Human migration denotes any movement by human from one locality to another. People who migrate are called migrants. Migration is determined by social, cultural and economic factors. Data on Migration was first recorded in the Census of India conducted in 1881 according to place of birth. Migration leads to the redistribution of the population within a country. It results in balanced distribution of people according to resources. The distribution of populations can be defined by local, regional, national, global and with different types of boundaries such as political, economic, and geographic. Rural-urban migration is one of the important factors contributing to the population growth of cities. A person who moved from one administrative area to another is a migrant. The point of departure is known as area of origin and the point of arrival is known as area of destination. An in-migrant/immigrant is a person who enters a migration defining area by crossing its boundary from some point outside the area but within the same country where as an out-migrant/emigrant is a person who departs from a migration defining area by crossing its boundary to a point outside it, but within the same country. Then the balance between in-migrants and out-migrants is called as net migrants. If migration occurs between birth and the time of the census or survey is known as life time migration. According to migration streams, Intra-district migrants are persons with last residence outside the place of enumeration but within the same district, Inter-district migrants are persons with last residence outside the district of enumeration but within the same state and Inter-state migrants are persons with last residence in India but beyond the state of enumeration. Here we used the 2001 & 2011 census data on sex and stream wise migrants of India to estimate the pattern of migration in the country. Sandhya Rani Mahapatro(2008) studied the emerging patterns in India and analyze that migrants are attracted to urban areas in recent times. By using NSS round data 2007/08 on migration and concluded that urbanization will change the nature of the economy and by comparing the NSSO findings with 2011 census results that there is an increasing mobility among females for education reflects social development. Alastair Smith (2003) observed that Markov chain Monte Carlo offers a powerful estimation tool. He analyses a simple linear model in the Bayesian framework and examine the properties of Markov chain for estimation. Mustofa Usman, Faiz A. M. Elfaki (2015) studied the markov chain model for migration in Indonesia for the years 1980-2010. They analyse the properties of transistion probability matrices, to find the stationary probabilities and to find the behavior of the mechanism of the migration with individual and combined data for the years 1980-2010. Our aim is to study the migrant's dispersal over rural and urban, to study the gender wise dispersal of migrants in rural and urban and to estimate the urbanization over migration effect of India using Markov chain model.

2. Materials Methods:

2.1 Markov process:

A discrete valued stochastic process is called a markov process. The markov property is that the probability distribution over the next observation depends only upon the current observation (Alaistair Smith,2003). A Markov Chain is a special type of stochastic process for projection which may be described as follows. Let X_i be the number of values at time t. The values of the variable X_t are called the states of the Markov Chain. At any given time t, when the current state X_t and all previous states $X_1,X_2, \ldots,X_{t-1}$of the process are known, the probabilities of all future states $X_j(j>t)$ depend only on the current state X_t and do not depend on the earlier states $X_1, X_2, \ldots, X_{t-1}$. Formally a Markov chain is a stochastic process such that for t = 1,2,… and for any possible sequence of states $X_1, X_2, \ldots, X_{t+1}$.

$$P\{X_{t+1} = x_{t+1} / X_1 = x_1, X_2 = x_2 \ldots X_t = x_t\} = P\{X_{t+1} = x_{t+1} / X_t = x_t\} \quad (1)$$

2.1.1 Markov Chain and transition probability matrix:

If the parameter space of a markov process is discrete then the markov process is called a markov chain. Let P be a (k x k)-matrix with elements P_{ij} (i, j = 1,2,…,k). A random process X_t with finite number of k possible states S = { $s_1, s_2 \ldots s_k$ }at any time the chain is called a finite Markov chain with transition matrix P. The Conditional probability P{ $X_{t+1} = s_j / X_t = s_i$} that the Markov chain will be in state s_j at time t+1 if it is in state s_j at time t is called a transition probability. That is for all t and all i, j = 1,2,…k we have,

$$P\{X_{t+1} = s_j / X_t = s_i\} = P_{ij} \text{ for } t = 1,2, \ldots \quad (2)$$

A convenient way to present the transition probabilities is through transition matrix. That is the elements of a transition matrix P are called transition probabilities. The transition matrix of the Markov chain is defined to be the k x k matrix p with elements P_{ij} is as follows,

$$P = \begin{pmatrix} P_{11} & P_{12} & \ldots & P_{1k} \\ P_{21} & P_{22} & \ldots & P_{2k} \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ P_{k1} & P_{k2} & \ldots & P_{kk} \end{pmatrix} \quad (3)$$

Every transition matrix satisfies, **$P_{ij} \geq 0$ for all i,j = 1,2,…k and $\sum P_{ij} = 1$ for i =1,2,..k** because if the chain is in state s_1 at a given observation time, then the sum of the probabilities that it will be in each of the $s_1, s_2 \ldots s_k$ at the next observation time must be 1. A square matrix for which all elements are non-negative and the sum of the elements in each row is 1 is called a stochastic matrix.

2.1.2 n- step Transition probability matrix:

Let P be the transition probability matrix of a finite Markov chain will elements P_{ij} (i,j = 0,1,2,…k-1). Then the n- step transition probabilities $P_{ij}^{(n)}$ are obtained as the elements of the matrix P^n. Let the row vector P^0= ($p_0^{(0)}, p_1^{(0)}, \ldots$) be a given vector of initial state occupation probabilities and let $P^{(n)}$= ($p_0^{(n)}, p_1^{(n)}, \ldots$) be the vector of state occupation probabilities at time t, the n-step transistion probabilities are given by,

$$P^{(n)} = P^{(0)} . P^{n-1} \ (n = 1,2,\ldots) \quad (4)$$

The components of $P^{(n)}$, (n = 0,1,…) are defined by $P_t^{(n)}$ = P [$X_1 = t_i$] , (n = 0,1,2,…). By the recurrence relation, we observe that for all i,j $P_t^{(n)} = \sum P_t^{(n-1)}$. P (n= 1,2,…). Since $p_t^{(n-1)}$ is a probability distribution and P_{ij} is non-negative and bounded by 1 for all i and j, the series on the right hand side of the equation is absolutely convergent. We may thus express in a matrix notation

$$P^{(n)} = P^{(n-1)} . P \quad (5)$$

2.1.3 Markov chain Model for migration :

Consider a Markov Chain with 5 time periods t_i = 1961, 1971,1981, 1991 and 2001 which all are census years of India. Let X_i be the percentage of migrants of each sex by migration streams in India at time t_i. Let x_1, x_2, x_3, x_4, and x_5 be the states of the Markov Chain.

x_1 = the percentage of migrants in migration stream by sex at the year 1961
x_2= the percentage of migrants in migration stream by sex at the year 1971
x_3= the percentage of migrants in migration stream by sex at the year 1981
x_4 = the percentage of migrants in migration stream by sex at the year 1991
x_5 = the percentage of migrants in migration stream by sex at the year 2001.
At a given time t, the current state may be the present years.

2.1.3.1 Condition for Markov chain Model:

The value of the current state depends only on the immediate previous time point and not on the past history.

$$\text{ie., } P\{x_{t+1} = x_{t+1} / x_1 = x_1, x_2 = x_2, \ldots x_t = x_t\} = P\{X_{t+1} = x_{t+1} / X_t = x_t\}$$

2.1.3.2 Transistion probability matrix for migration:

$$P = \begin{cases} P^M = [\, P_{ij}^M \,]\ (i, j = 1,2) \text{ which is the probability of male migrants between streams.} \\ P^F = [\, P_{ij}^F \,]\ (i, j = 1,2) \text{ which is the probability of female migrants between streams.} \end{cases} \quad (6)$$

Let $P_{t_i}^M$= ($P_{t_i}^{RM}, P_{t_i}^{UM}$) be a given vector of initial state probabilities for migration of males at time t_i, which is the census year.

$$P_{t_i}^M = \begin{cases} P_{t_i}^{RM} \text{ - Probability of male migrant is from rural population.} \\ P_{t_i}^{UM} \text{ - Probability of male migrant is from urban population.} \end{cases} \quad (7)$$

Similarly, $P_{t_i}^{F} = (P_{t_i}^{RF}, P_{t_i}^{UF})$ be a given vector of initial state probabilities for migration of females at time $t_{i.,}$ which is the census year.

$$P_{t_i}^{M} = \begin{cases} P_{t_i}^{RM} \text{ - Probability of female migrant is from rural population.} \\ P_{t_i}^{UM} \text{ - Probability of female migrant is from urban population.} \end{cases} \quad (8)$$

where,

$$P_{t_i}^{RM} = \frac{Number\ of\ Rural\ male\ population}{Total\ Rural\ population} \text{ X } \frac{Number\ of\ Rural\ male\ migrants}{Total\ Rural\ migrants}$$

$$P_{t_i}^{UM} = \frac{Number\ of\ urban\ male\ population}{Total\ urban\ population} \text{ X } \frac{Number\ of\ urban\ male\ migrants}{Total\ urban\ migrants}$$

$$P_{t_i}^{RF} = \frac{Number\ of\ Rural\ female\ population}{Total\ Rural\ population} \text{ X } \frac{Number\ of\ Rural\ female\ migrants}{Total\ Rural\ migrants}$$

$$P_{t_i}^{UF} = \frac{Number\ of\ Rural\ female\ population}{Total\ Rural\ population} \text{ X } \frac{Number\ of\ Rural\ female\ migrants}{Total\ Rural\ migrants}$$

The above proportions indicate the distribution of migrants intensity in rural and urban at time $t_{i.,}$ then for future time points, $P_{t_{i+1}} = P_{t_i} . P$ which is convergent for all census years.

For example, $P_{t_i}^{RM}$ = (225.3/439.1)x(0.6) = 0.30785, and $P_{t_i}^{UM}$ = (58.72/109.1)x(0.4) = 0.21528, t_i is 1971, then the initial state probability is given by, $P_{t_i}^{M} = (P_{t_i}^{RM}, P_{t_i}^{UM})$ = (0.30785, 0.21528).

Let $P = P^{M} = \begin{pmatrix} 0.6729 & 0.3270 \\ 0.3170 & 0.6829 \end{pmatrix}$ be the transition matrix for the year 1971 by stream then,

$P_{t_{i+1}} = P_{t_i} . P$ is the projection time points.

$$P_{1981} = (0.3078, 0.2152)\begin{pmatrix} 0.6729 & 0.3270 \\ 0.3170 & 0.6829 \end{pmatrix} = (0.2754, 0.2477)$$

$P_{1991} = (0.2754,0.2477)\begin{pmatrix} 0.6729 & 0.3270 \\ 0.3170 & 0.6829 \end{pmatrix}$ = (0.2639,0.2592) and so on for t_i = 2001,2011,2021…

2.2. Test for goodness of fit:

The test for goodness of the model is given by Chi-Square distribution. For testing, H_o : The model is a good fit, we use chi square distribution at α =5% level of significance with (n-1) degrees of freedom. The test statistic for testing H_0 is given by,

$$\chi^2 = \frac{(O-E)^2}{E} \sim \chi^2_{(n-1)d.f} \quad (9)$$

If the calculated value is less than the table value then H_0 is accepted which infers the model is a good fit. Otherwise H_0 is rejected.

2.3. Source of data:

The secondary data is collected from census of India for census years from 1961 to 2001.

3. Statistical Analysis and result discussions:

The following table 3.1 shows the data on percentage distribution of migrants by sex in migration streams which is observed from census of India.

Table 3.1 : Percentage Distribution of Migrants in Different Migration Streams by Sex

Sex	Year	Rural to urban	Urban to urban	Urban to rural	Rural to rural	Total migrants
Female	1961	9.7	5.8	3.2	81.3	100
	1971	10.5	6.7	5.1	77.7	100
	1981	12.5	8.7	5.5	73.3	100
	1991	13.5	8.8	5.5	72.2	100
	2001	13.6	9.7	5.6	71.1	100
	1961	25.7	13	4.6	56.7	100
	1971	26	14	6.5	53.5	100

Male	1981	30	17.4	7	45.6	100
	1991	31.6	17.8	7.2	43.4	100
	2001	27.1	18.3	8.6	46	100

The following table shows the sex wise population of India (in millions) since 1901 according rural and urban distribution.

Table 3.2 : Sex wise Rural-Urban distribution of Population since 1901 of India (in Millions)

Year	Population								
	Rural			Urban			Total		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
1901	107.48	105.22	212.70	13.46	12.24	25.70	120.94	117.46	238.40
1911	114.48	111.62	226.10	13.89	12.11	26.00	128.37	123.73	252.10
1921	113.30	109.90	223.20	15.22	12.88	28.10	128.52	122.78	251.30
1931	124.87	120.63	245.50	18.23	15.27	33.50	143.10	135.90	279.00
1941	139.44	134.56	274.00	24.14	20.06	44.20	163.58	154.62	318.20
1951	152.01	146.69	298.70	33.55	28.85	62.40	185.56	175.54	361.10
1961	183.55	176.75	360.30	42.76	36.14	78.90	226.31	212.89	439.20
1971	225.30	213.80	439.10	58.72	50.38	109.10	284.02	264.18	548.20
1981	269.35	256.15	525.50	84.99	74.71	159.70	354.34	330.86	685.20
1991	324.23	304.44	628.67	115.52	102.71	218.23	439.75	407.15	846.90
2001	381.14	360.52	741.66	150.12	135.22	285.34	531.26	495.74	1027.00
2011									

The calculated year wise probability distribution of migrants by sex in different streams by equation (6) is given in table 3.3 which is as follows.

Table 3.3 : Probability distribution of migrants in different migration streams by sex

Year	Sex / Streams	MALES		FEMALES	
		Rural	Urban	Rural	Urban
1961	**Rural**	0.6881	0.3118	0.8934	0.1065
	Urban	0.2613	0.7386	0.3555	0.6444
1971	**Rural**	0.6729	0.3270	0.8809	0.1190
	Urban	0.3170	0.6829	0.4322	0.5677
1981	**Rural**	0.6031	0.3968	0.8543	0.1456
	Urban	0.2868	0.7131	0.3873	0.6126
1991	**Rural**	0.5786	0.4213	0.8424	0.1575
	Urban	0.288	0.712	0.3846	0.6153
2001	**Rural**	0.6292	0.3707	0.8394	0.1605
	Urban	0.3197	0.6802	0.3660	0.6339

The below table 3.4 shows the initial state probabilities (bold case) of male migrants which is computed from equation (7) and projected migrant distribution of males by taking t_i as base year using markov chain.

Table 3.4: Year wise projected distribution of migrants for males in India										
Year	**1961**		**1971**		**1981**		**1991**		**2001**	
	Rural	**Urban**	**Rural**	**Urban**	**Rural**	**Urban**	**Rural**	**Urban**	**Rural**	**Urban**
1961	**0.3108**	**0.2114**								
1971	0.2691	0.2530	**0.3079**	**0.2153**						
1981	0.2513	0.2708	0.2754	0.2477	**0.2665**	**0.2501**				
1991	0.2437	0.2784	0.2639	0.2592	0.2325	0.2841	**0.2621**	**0.2578**		
2001	0.2405	0.2817	0.2598	0.2634	0.2218	0.2949	0.2259	0.2940	**0.2775**	**0.2367**
2011	0.2391	0.2830	0.2583	0.2648	0.2184	0.2983	0.2154	0.3045	0.2503	0.2639
2021	0.2385	0.2836	0.2578	0.2653	0.2173	0.2994	0.2123	0.3076	0.2419	0.2724
2031	0.2382	0.2839	0.2576	0.2655	0.2169	0.2997	0.2114	0.3084	0.2393	0.2750

The above table concludes the distributions of rural males have decreasing trend. For urban males the distribution has increasing trend.

Figure (1): Rural- Urban distribution of projected migrant probabilities of Males in India

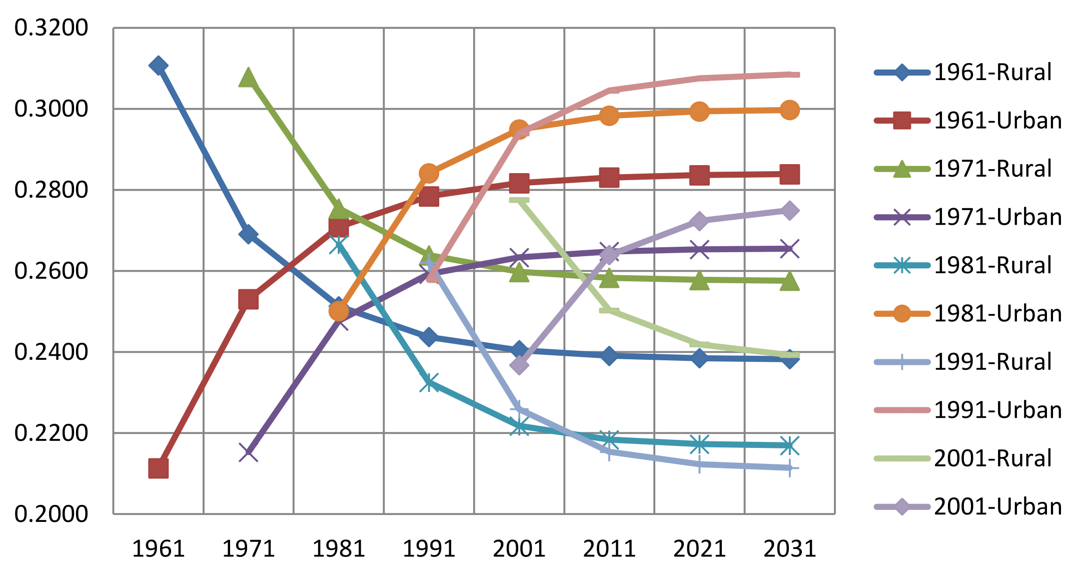


Figure (1) shows the stationary probabilities of male migrants according to Rural and Urban which are computed from markov chain model results in convergence.

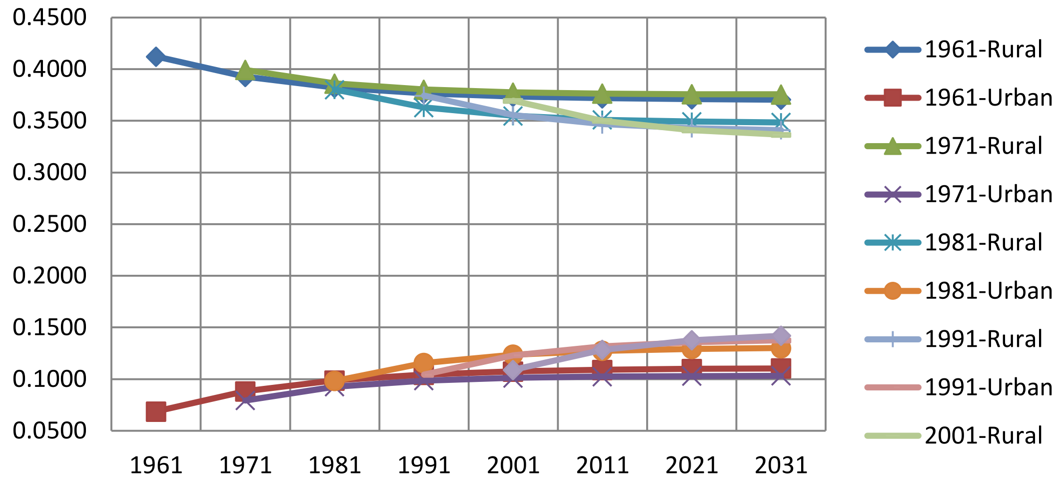
Similarly, The below table shows the initial state probabilities of female migrants (bold case) which is computed from equation (8) and projected migrant distribution of females by taking t_i as base year.

Table 3.5: Year wise projected distribution of migrants for females in India										
Year	**1961**		**1971**		**1981**		**1991**		**2001**	
	Rural	**Urban**	**Rural**	**Urban**	**Rural**	**Urban**	**Rural**	**Urban**	**Rural**	**Urban**
1961	**0.4121**	**0.0687**								
1971	0.3926	0.0882	**0.3993**	**0.0794**						
1981	0.3821	0.0987	0.3861	0.0926	**0.3802**	**0.0982**				
1991	0.3765	0.1043	0.3801	0.0986	0.3629	0.1156	**0.3743**	**0.1043**		
2001	0.3734	0.1074	0.3775	0.1012	0.3548	0.1237	0.3554	0.1231	**0.3694**	**0.1090**

2011	0.3718	0.1090	0.3763	0.1024	0.3510	0.1275	0.3468	0.1318	0.3500	0.1284
2021	0.3709	0.1099	0.3757	0.1029	0.3492	0.1292	0.3428	0.1357	0.3408	0.1376
2031	0.3704	0.1103	0.3755	0.1032	0.3484	0.1301	0.3410	0.1375	0.3365	0.1420

Table 3.5 concludes the distributions of rural females have decreasing trend. For urban females the distribution has increasing trend. Also Figure 2 shows the stationary probabilities of female migrants according to Rural and Urban which are computed from markov chain model results in convergence.

Figure (2): Rural- Urban distribution of projected migrant probabilities of females in India



4. Testing for goodness of fit:

Consider the hypothesis, H_o **:** The fit is good for Rural and Urban males in India. From table 3.4 the goodness of fit is estimated by chi-square statistic.

Table 4.1 Chi –square table for rural and Urban males in India						
Year	**Rural Males**			**Urban Males**		
	Observed(O)	**Expected(E)**	**(O-E)^2/E**	**Observed(O)**	**Expected(E)**	**(O-E)^2/E**
1971	31	27	0.592593	22	25	0.36
1981	27	28	0.035714	25	25	0
1991	26	23	0.391304	26	28	0.142857
2001	28	28	0	24	29	0.862069
	Calculated χ^2 value		**1.019611**	Calculated χ^2 value		**1.364926**

At α= 5% level of significance for testing H_o, the calculated chi-square values for rural males and urban males is less than the chi-square table value (7.815) for n-1=4-1=3 d.f , our hypothesis is accepted which concludes that the model is a good fit.

Consider the hypothesis, H_o **:** The fit is good for Rural and Urban females in India. From table 3.5 the goodness of fit is estimated by chi-square statistic.

Table 4.2 Chi –square table for rural and Urban females in India						
Year	**Rural Females**			**Urban Females**		
	Observed(O)	**Expected(E)**	**(O-E)^2/E**	**Observed(O)**	**Expected(E)**	**(O-E)^2/E**
1971	40	39	0.025641	7	9	0.444444
1981	38	39	0.025641	10	8	0.5

1991	38	36	0.111111	10	13	0.692308
2001	37	36	0.027778	11	12	0.083333
	Calculated χ^2 value		**0.190171**	Calculated χ^2 value		**1.720085**

At α= 5% level of significance for testing H_o, the calculated chi-square values for rural females and urban females is less than the chi-square table value (7.815) for n-1=4-1= 3 d.f , our hypothesis is accepted which concludes that the model is a good fit.

5. Conclusion:

In India, migration influenced in shaping our culture as well as diversity of population. In the present scenario, migration is controlled by 'pull' and 'push' factors. In the long run, the density of the populations in rural and urban areas will be almost equal. The study reveals that the projected probabilities of migration after 20 years are more or less stable in India for respective base years. This is specifically true in the case of male migrants. The probability of male migration to rural areas has a downward trend for both sex wise and stream wise cases. This reveals that rapid urbanization will be the effect of male migration and female migration. This rapid urbanization has reduced the size of agricultural land holdings, average land holding size etc., which increases pressure on land and no longer able to support a family. This pressure on land leads to the growth of slums. The drawbacks of downward trend reveal the changes in behavior and attitude of the female migrants which can be rectified by improving education and employment in rural areas. Because females in India have always suffered from lower status, early marriages, lower literacy, poor nutrition and higher fertility and mortality levels during reproductive age (Anjana Mazumdar,2011) whereas for males in rural suffers by less opportunities of employment, low level income, lack of education and training facilities, lack of medical facilities etc., which can be rectified by developing rural development programmes and hence from the above cases, there is an indication to urbanization. According to the National Policy on population, urbanization leads to improving the quality of life and economic growth. It is because urbanization provides more technological challenges to satisfy the needs of the people and hence more opportunities for the human talents. Hence the pattern of migration for both sex and stream wise will intensify the growth of the nation in future.

References:

[1]. Alastair Smith. 2003, Markov chain monte carlo simulation made simple, Newyork University, Newyork.
[2]. Andrei Rogers, A Markovian Analysis of Migration differentials, University of Kalifornia.
[3]. Asis Kumar Chattopadhyay and Anuj Kumar Saha. 2012, Demography Techniques and Analysis, Viva Books Private Ltd., Chennai.
[4]. Censusindia.gov.in, Office of the Registrar General & Census Commissioner, Ministry of Home Affairs, India.
[5]. Mustofa Usman, Faiz A. M. Elfaki, Yusuf. S. Barusman, Widiarti and Ayu sofya. 2015, Markov chain analysis and mechanism of migration in Indonesia in the period of 1980-2010, ARPN Journal of Engineering and applied sciences, ISSN 1819-6608. Vol.10, No.22, Dec 2015, Lampung University, Indonesia.
[6]. Sandhya Rani Mahapatro 2008, The changing pattern of Internal migration in India, Issues and challenges, ISEC, Bangalore.
[7]. Suddhendu Biswas. 1988, Stochastic Processes in Demography and Applications, Wiley Eastern Limited, ISBN 81-224-0061-2., NewDelhi.
[8]. Reuven Y. Rubinstein. 1981, Simulation and the Monte carlo method, John Wiley and Sons, Newyork.
[9]. Tom Wilson. 2011, A review of Sub- regional population projection methods, University of Queensland, Australia.