The Nepali Math. Sc. Report
Vol. 31, No. 1 & 2, 2011

# Markov Chain Model to Describe the Distribution of Intergenerational Occupational Mobility

KNS YADAVA[1] AND TR ARYAL[2]

[1]Department of Statistics
Banaras Hindu University, Varanasi, India
e-mail: knsy@bhu.ernet.in

[2]Central Department of Statistics
Tribhuvan University, Kirtipur Kathmandu, Nepal
e-mail: traryal@gmail.com & traryal@rediffmail.com

**Abstract:** This paper attempts to apply a Markov chain model to study intergenerational occupational mobility distribution among the residents of Palpa and Rupandehi districts. Markov chain model was found to be the appropriate model to describe the distribution of intergenerational occupational mobility. The attraction of sons was more toward non-manual and non-agricultural occupations as compared to their fathers. Indeed, intergenerational occupational mobility was found moderately high for both the generations. Findings may help planners and policy-makers in designing proper policies especially in integrated rural development program.

## 1. Introduction

The dynamic structure of social phenomena is linked with the movement of the people across social, economic or occupational categories. In this line several studies on occupational mobility have been carried out by scholars [1,2,3,4,5,6,7]. Models for mobility involved probabilistic terms which are useful for the future prediction and also for assessing the likely error of these predictions [8]. The need of the measurement of social mobility was first realized in connection with the empirical research [3]. Prais [6] was probably the first researcher to apply Markov chain theory to social mobility followed by the measurement of occupational mobility based on the semi-Markov process [9].

Most of the mobility measures are developed based on the elements of transition matrix, which may be recognized by the transition probability matrix. A number of researchers have been studied the intergenerational occupational mobility in different societies and communities [1,2,5,7,10,11]. Markov chains or mixtures of Markov chains, like mover-stayer model, have been commonly used in social sciences to study various forms of dynamic behavior of human being including occupational mobility [1,2,4,5,12].

This paper attempts to apply a Markov chain model to study the distribution of intergenerational occupational mobility among the residents of Palpa and Rupandehi districts. In brief, Markov chain model is given below.

## 2. Markov Chain Model

Markov chain model developed by Sampson [12] is applied to study the inter-generational occupational mobility pattern. Let $Q_u = (p_{ij})$ be a k x k unrestricted transition matrix, where $p_{ij}$ is the probability that a process is in state j (j=1,2,3 . . .k), the existing period, given that it was in state i(i=1,2,3 . . .k) in the initial period, and k is the occupational group, hence forth known as state of a Markov chain. The class of restricted transition matrices is given as:

$$Q_r = \theta + (I - \theta)vp^* \; ;$$

$$\text{or} \quad p_{ij} = \theta_i \delta_{ij} + (1-\theta_i)p_j^* \tag{1}$$

where, I is the identity matrix, and $\delta_{ij}$ is the kronecker delta, v be the k x 1 vector of ones, and $p^*=(p_j^*)$ be a 1x k vector of probabilities, such that $p^*v=1$, and $\theta=\text{diag}(\theta_i)$ be an k x k diagonal matrix with $0\leq\theta\leq 1$. $(1-\theta_i)$ represents to the probability that a son would look for a different occupation that of his father and $\theta_i$ is the probability that a son would look for the same occupation that of his father.

Now, pre-multiplying equation (1) by p, we have,

$$p^* = p(I-\theta) / p(I-\theta)v \; ,$$

or $p_j^* = p_j(1-\theta_j) / \sum p_k (1-\theta_k)$ (2)

On substituting equation (2) into equation (1), we get

$$Q = \theta + \frac{(I - \theta)vp(I - \theta)}{p(I - \theta)v},$$

$$\text{or} \quad p_{ij} = \theta_i \partial_{ij} + \frac{(1-\theta_i)(1-\theta_j)p_j}{\sum_k p_k(1-\theta_k)} \tag{3}$$

Let $p=(p_j)$ be the equilibrium distribution of Q, so that, $pQ = p$. In general, $p^* \neq p$, except for the condition that all the values of $\theta_i$'s are identical.

On solving equation (1) by using $pQ=p$, we get the equilibrium distribution as given below,

$$p = p^*(I-\theta)^{-1}/p^*(I-\theta)^{-1}v,$$

$$\text{or} \quad p_j = p_j^*(1-\theta_j)^{-1}/\sum_k p_k^*(1-\theta_k)^{-1} \tag{4}$$

The occupation of sons will have the probability $p_j^*$ of being in occupational category j, which is independent of occupational category i. Furthermore, p and v be the two eigen vectors, which are associated with the eigen values 1 to the left and v to the right respectively. It was also shown that the remaining eigen values are the root of the function $f(\lambda) = p^*(\theta - \lambda I)^{-1}v$. where $\lambda$ is the scalar quantity having only one eigen value lying between $\theta_i$ and $\theta_{i+1}$ when $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_s$. The right and left eigen vectors corresponding to the eigen values $\lambda$ are treated as $p^*(\lambda I-\theta)^{-1}$ and $(\lambda I-\theta)^{-1}(I-\theta)v$ respectively.

Two parameters $\theta$ and $p^*$ are to be estimated by maximum likelihood (ML) method. In a Markov chain model, let N be the independent realization of length T+1 period for t= 0, 1, 2, ..., T, and log-likelihood function is given by

$$l = \sum n_{ij} \ln(p_{ij}) \tag{5}$$

where $n_{ij}$ is the number of transitions in the sample from state i to a state j (i.e. father's occupational state i to the son's occupational state j). The unrestricted ML estimates are given as,

$$P_{ij} = \frac{n_{ij}}{n_i} \tag{6}$$

It is observed that,

$$n_i = \sum_j n_{ij},\ \ddot{n}_i = \sum_j n_{ji} \text{ and } n = \sum \sum_j n_{ij} = \sum_i n_i = \sum_i \ddot{n}_i$$

where $n_i$ is the number of the occurrence of i for the initial period of state and $\ddot{n}_i$ is the number of the occurrences of i for the existing period of state.

So the 1xk probability vector is $\overline{p} = (n_i / n)$ and $\ddot{p} = (\ddot{n}_i/n)$ satisfies the following equation,

$$\ddot{p} = \overline{p}\, Q_u \tag{7}$$

where $Q_u = (p_{ij})$ is the unrestricted ML transition matrix.

Similarly, restricted ML estimates are obtained by substituting equation (1) in equation (5), and we obtain the restricted log-likelihood function as given below,

$$l = \sum\{n_{ii} \ln[\theta_i + (1-\theta_i)p^*] + (\ddot{n}_{.i} - n_{ii}) \ln(p_i^*) + (n_i - n_{ii})\ln(1-\theta_i)\} \tag{8}$$

Maximizing equation (8) with respect to $\theta_i$ and p* and some simplification was made by Aryal [1,2], we get,

$$\hat{\theta}_{it} = \hat{\theta}_{it-1} - G_i a_i + \frac{a_i \sum_k \overline{p}_k a_k b_k G_k}{1 + \sum_k \overline{p}_k a_k b_k} \tag{9}$$

$$a_i(\theta) = [1 + (2\theta_i \overline{p}_i - \ddot{p}_i - \overline{p}_i)/\{\overline{p}(I-\theta)v\}]^{-1} \text{ and}$$

$$b_i(\theta\theta = (1-\theta_i)(\ddot{p}_i - \overline{p}_i\theta_i)/[\overline{p}(I-\theta)v]^2.$$

Once we obtain $\hat{\theta}_i$, $\hat{p}_i^*$ is easily estimated by the equation

$$\hat{p}_i * = 1 - \left(1 - \frac{n_{ii}}{n_i}\right) / \left(1 - \hat{\theta}_i\right) \tag{10}$$

## 3. Application

The model is tested with the real sets of data from a sample survey of Palpa and Rupandehi districts. A total of 811 households were surveyed. The information on occupation was collected from each household. This paper deals with a sample of 777 fathers and sons of the first two generations and 303 fathers (sons) and sons (grandsons) of last two generations. The occupational categories are noted as: 1=Agricultural Laborer (Land less), 2=Non- Agricultural Laborer, 3=Agriculturist (landowner), 4=Contractual workers and Workers in Abroad, 5=Business/Trade and household industry, and 6=Professional, Administrative and Govt./Pvt. Service. The details of data and measurements of variables are found in Aryal [13,14,15]

Table 1: Estimated transition probability matrix (first two generations)

| Occupational categories | | Son | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| Father | 1 | .3482 | .1696 | .1071 | .1339 | .0893 | .1518 |
| | 2 | .2299 | .4138 | .0690 | .0690 | .0920 | .1264 |
| | 3 | .0862 | .0366 | .3190 | .1638 | .1228 | .2716 |
| | 4 | .0364 | .0182 | .1636 | .4000 | .1273 | .1818 |
| | 5 | .0385 | .0000 | .1538 | .0000 | .4231 | .3846 |
| | 6 | .0000 | .0303 | .3333 | .0606 | .2121 | .3636 |

Table 1 shows the occupational distribution of sons by occupation of their fathers in the first two generations (from father to son). The estimated transition probability matrices are also shown in the respective table for the first two generations. The occupational categories 1 to 6 represent the different states of a Markov chain model under consideration and the mobility either forward or backward is of one step (father and son).

Table 2: Estimated transition probability matrix (last two generation)

| | | Son | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| Father | 1 | .2826 | .1087 | .1522 | .2174 | .0217 | .2174 |
| | 2 | .1905 | .3810 | .0476 | .2619 | .0000 | .1190 |
| | 3 | .0400 | .0400 | .2800 | .2000 | .1200 | .3200 |
| | 4 | .0750 | .0250 | .1750 | .4750 | .0750 | .1750 |
| | 5 | .0000 | .0000 | .1333 | .0333 | .5333 | .3000 |
| | 6 | .0000 | .0889 | .1111 | .0889 | .0889 | .6222 |

Table 2 presents the occupational distribution of the last two generations (from sons to grandsons). The estimated transitional probability matrices are also displayed in this table for the last two generations.

Table 3: Estimated parameters (first two generations)

| Occupational categories | θ | p* | p |
|---|---|---|---|
| 1 | .2601 | .1158 | .1250 |
| 2 | .3801 | .0527 | .0680 |
| 3 | .1850 | .1599 | .1567 |
| 4 | .2634 | .1803 | .1955 |
| 5 | .3159 | .1523 | .1778 |
| 6 | .0231 | .3390 | .2770 |
| -2logL(Null) | | | 1227.086 |
| -2logL(Model) | | | 1225.783 |
| Model(LR) chi-square | | | 2.61 |
| Degrees of freedom | | | 16 |
| Critical values(5% level) | | | 26.30 |

The estimated values of parameters θ and p* are given in Tables 3 and 4 for both the successive generations. An estimate of θ explains that the son follow the same occupation under taken by his father whereas p* shows the chance of getting an occupation different from his father's occupation. However, when $\theta_j$'s are identical, the estimated value of parameter p* gives a state of equilibrium. This is considered as the persistence class of Markov chains [1,2,5] whereas when $\theta_j$s are equal to zero, the process exhibits an inter-temporal independence [1,5].

Thus, $\theta_j$ determines the extent to which state i (the initial structure) influences the next period of state i.e. the existing structure. Similarly, when $\theta_j$=0, then the initial state has no influence on the existing period of state. Moreover, when $\theta_j$=1, and obviously, $p_{ii}$=1, then i becomes an absorbing state of Markov chain, the process in that state remains there forever, i.e. immobile [1,2].

Table 4: Estimated parameters (last two generations)

| Occupational categories | θ | p* | p |
|---|---|---|---|
| 1 | .2358 | .0661 | .0560 |
| 2 | .3510 | .0502 | .0501 |
| 3 | .1211 | .1930 | .1432 |
| 4 | .3032 | .2642 | .2474 |
| 5 | .4991 | .0731 | .0950 |
| 6 | .4330 | .3553 | .4089 |
| -2logL(Null) | 459.10 | | |
| -2logL(Model) | 423.65 | | |
| Model chi-square | 70.80 | | |
| Degrees of freedom | 15 | | |
| Critical values (5% level) | 25.00 | | |

Thus, a person in the occupational category of agricultural laborer, for example, has had an estimate (1-0.2601) =0.7399 probability of getting a different occupation of son from his father whereas the son has a chance of 0.0527 of joining an occupation in non- agricultural laborer (Table 3). Similarly, a person in the occupational category non- agricultural laborer, has had an estimate (1-.3510)=. 6490 probability of getting a different occupation than that of his father whereas the son has a chance of .3553 of joining an occupation in professional, administrative and government and private service for the last two generations (Table 4) and so on. For the other elements of the estimates, a similar interpretation can be given.

The observed log-likelihood ($L_0$), is -1227.086 and expected log-likelihood ($L_a$), is -1225.783 (Table 3). Thus, the likelihood ratio (LR) test statistic for $H_0$: $Q=\theta+(I-\theta)vp^*$ is $-2(L_0 - L_a)=2.61$ at the 16 degrees of freedom which is accepted as the tabulated $\chi^2(16) = 26.30$ at the 5% level of significance for the first type of data set i.e. first two generations (fathers to sons). An insignificant value of chi-square suggests that the Markov chain model fitted the data very well. But from Table 7.4, the likelihood ratio (LR) test statistic for $H_0 : Q=\theta+(I-\theta)vp^*$ is =70.80 at the 15 degrees of freedom and comparing with $\chi^2(15) = 25$ (at the 5% level), the null hypothesis is rejected for the second set of the data i.e. the last two generations (sons to grandsons).

The main diagonal elements of the transition probability matrix show the inheritance (son's occupation is same as that of their father's) occupational situations among the six occupational categories for both the generations (Tables 1 and 2). Occupational inheritance was considerably similar (.3482, .4138, .3190, .4000, .4231 and .3636) but low in nature among all the occupational categories (Table 1) whereas occupational inheritance was fluctuating in behavior (.2826, .3810, .2800, .4750, .5333 and .6222) among the six occupational categories (Table 2). Occupational inheritance was low for the first three occupational categories i.e. agricultural laborer, non-agricultural laborer and agriculturist (landowner) whereas it was high for the last three occupational categories i.e. professional, administrative & government and private services; contractual workers and workers abroad for the last two generations.

It is found that the attraction of sons was more toward the non-manual and non-agricultural occupations and they were highly mobile towards these occupational categories. It may be due to changed occupational status of the rural people because of many influencing factors such as spread of education, industrial development, constant land resources for cultivation, rise of the population, climatic conditions, rural-urban migration, dissolution of the households and government policies. Moderate amount of mobility also occurred towards the manual and agricultural occupations. This may be due to some specific characteristics of individual who need for joining non-manual and non-agricultural occupations.

## 4. Conclusions

Markov chain model has been applied to describe the distribution of intergenerational occupational mobility among the residents of Palpa and Rupandehi districts. It was found that the attraction of sons was more toward the non-manual and non-agricultural occupations. The intergenerational occupational mobility was found moderately high for both the generations. Findings may have a number of policy implications particularly for the developing countries, like Nepal. This may also help the planners and policy makers in designing policies such as integrated rural development program, etc. to be adopted for specific region.

## REFERENCES

[1] Aryal, T.R. 2006. Inter-generational occupational mobility among residents of rural Nepal, *Journal of Institue of Science and Technology*, Vol. 14, pp. 123-135.

[2] Aryal, T.R. 2002. *Some demographic models and their applications with reference to Nepal*, Ph.D. thesis, Department of Statistics, BHU, India.

[3] Chattopadhyay, A.K. and K. Baidya. 1999. Social mobility among residents of Calcutta, *Demography India,* vol. 18(2):215-224.

[4] Krishna, R. and B.K. Pattnaik. 1997. Occupational mobility in urban community", *Demography India,* vol. 26(2):207-227.

[5] Yadava, K.N.S. and T.R. Aryal. 2001. A model describing the pattern of inter-generational occupational mobility in rural Nepal, *Journal of Population and Social Studies*, vol. 9(2):35-53.

[6] Prais, S. J. 1955. Measuring social mobility, *Journal of the Statistical Society America,* vol. 118:56-66.

[7] Varma, G.V. and P.H. Rayappa. 1997. Occupational classification systems and intergenerational occupational mobility studies: some methodological issues, *Demography India*, vol. 26(2):229-240.

[8] Bartholomew, D.J. 1982. *Stochastic models for social processes*, 3rd ed., Chichester John Wiley and Sons.

[9] Bartholomew, D. J. 1967. *Stochastic models for social processes*, New York, John Wiley and Sons.

[10] Singh, V.S. 1992. A statistical study on some aspect of human mobility, Ph.D. thesis, Department of Statistics, BHU, India.

[11] Ali, A.F.I. 1987. Occupational mobility among the Muslim castes in a Bangladesh village, *Asian Profile*, vol. 15(4):357-367.

[12] Sampson, M. 1990. A Markov chain model for unskilled workers and the highly mobile, *Journal of the American Statistical Association*, vol. 85(409):178-180.

[13] Aryal, TR, 2010. A new Technique to construct the female marriage life-table, *Nepali Mathematical Sciences Report,* Vol. 30 (1&2) pp. 111-122.

[14] Aryal, TR. 2009. Estimation of the Duration of Post-partum Amenorrhea among Nepalese Mothers: Some Indirect Techniques, Nepal Journal of Science and Technology, Vol. 10, pp. 213-218.

[15] Aryal, TR. 2008. An indirect technique to estimate the duration of post-partum amenorrhea, Nepal Journal of Science and Technology, 8, 137-142.

□□□