JOURNAL OF POPULATION AND SOCIAL STUDIES Volume 14 Number 2 January 2006

The Distribution of the Number of Migrants at the Household Level

Kamlesh Kr. Shukla *

K.N.S. Yadava *

Introduction

Usually migration has been studied taking aggregate variation in relation to various social, demographic and spatial characteristics based on data at district state or nations as a whole. Micro-level studies of migration, though few in numbers have been under taken at community, village, household and individual levels, depending upon the objective and availability of data.

In developing countries, particularly in India where about more than seventy percent people still live in villages, migration from rural areas has become a major subject of interest for social scientists as well for planners. Generally in rural area, the occurrence of migration from household can be classified as (i) the households from where adult male members aged 15 years and above migrate alone leaving wives and children at home, (ii) the household where male members migrate with their wives and children, and (iii) the households where both types of migration as mentioned above take place.

Different types of mathematical models have been used to represent the observed phenomenon in a concise form different group of households as mentioned above. A study of migration through a model has greater importance because it explains the pattern, trend and volume of migration. Distribution of households according to the number of male migrants (aged fifteen years and above) has been fitted by a negative

---

* Department of Statistics, Banaras Hindu University-221005, India. E-mail: shukla22@yahoo.com; knsyadv@sify.com

binomial distribution (Singh and Yadava, 1981). A good number of studies took place after this work and several models have been purposed to study the pattern of rural male ($\geq$15 years) out migration (Hossain, 2000; Iwunor, 1995; Singh, 1992; Sharma, 1987; 1985; Yadava ,Tripathi and Singh, 1994). However, these models were not appropriate to fit the distribution of total number of migrants (including wife and children) consequently, several attempts have also been made to describe the distribution of households according to the total number of migrants under different assumptions (Kushwaha, 1992; Singh, 1985; 1990; Yadava and Yadava, 1988).

In the present paper some probability models as discussed above are used with a new set of data on the number of migrants at the household level particularly to test the suitability of models over time. The merits and demerits of the models applied have also been discussed.

Models

1. Distribution of households according to number of male migrants aged 15 years and above.

Looking as the nature of data two simplest distributions proposed by Sharma (1988) are utilized to describe the distribution of households according to the number of migrants (aged greater than 15years and above). These are

(I) Inflated geometric distribution, and

(II) Inflated logarithmic series distribution.

These distributions have been proposed on the basis of the following assumptions:

(a) Occurrence of a migrant from a household at the survey point may be with probability α and hence with probability (1-α) there is no migration.

(b) The probability of k males migrating from a household is more than the probability of (k+1) members migrating from a household (for k = 1, 2, 3…)

Let x be the random number of rural male ($\geq$ 15 years) out-migrants from a household, then from assumptions (a) and (b), the probability mass function (P_x) is as follows:

Inflated geometric : $P[x = 0] = P_0 = 1 - \alpha; \quad k = 0$ (1)

$$P[x = k] = P_k = \alpha(q^{k-1}.p), \ q = 1 - p > 0; \ (k > 0) \quad (2)$$

Inflated logarithmic: $P[x = 0] = P_o = 1 - \alpha; \quad (k = 0)$ (3)

$$P[x = k] = P_k = \alpha\left(\frac{-\theta^k}{K\ln(1-\theta)}\right), \ \theta > 0; \ (k > 0) \quad (4)$$

Estimation of the Parameters

Inflated geometric distribution involves two parameters α and p. Suppose N households are observed at random and Ni denotes the number of migrants from the i^{th} household, then

$$E(x) = \sum_{k=0}^{\omega} kp_k = \frac{\alpha}{p}$$

$$\text{i.e. } \overline{X} = \frac{\alpha}{p} \quad (5)$$

Where $\overline{X}$ is the sample mean i.e. average number of male out migrants per household? Let N_o denote the number of households with no out-migrant, then an estimate of α is,

$$\hat{\alpha} = \frac{N - N_o}{N} \quad (6)$$

Then from (5) an estimate of p is

$$\hat{p} = \frac{\hat{\alpha}}{\overline{x}} \quad (7)$$

Similarly,

$$\text{var (x)} = \alpha\sum_{k=o}^{\infty} k^2 P_k - \left(\frac{\alpha}{p}\right)^2$$

$$= \alpha(1 + q - \alpha)/p^2$$

Inflated logarithmic distribution also consists of two parameters α and θ. Suppose N households are observed at random and Ni denotes the number of migrants from the ith household, then

$$E\,(x) = \sum_{k=0}^{\infty} kp_k = \frac{-\alpha\theta}{(1-\theta)\ln(1-\theta)}$$

$$\text{i.e. } \bar{x} = \frac{-\alpha\theta}{(1-\theta)\text{In}(1-\theta)} = \sum \frac{iNi}{N} \qquad (8)$$

Where $\overline{X}$ is the sample mean, i.e. average number of out migrants per household. Let N_o denote the number of households with no out migrants; we have,

$$\hat{\alpha} = \frac{N - No}{N} \qquad (9)$$

and then from (8), an estimate of θ, after some iteration, can be found as:

$$\frac{\hat{\theta}}{(1-\hat{\theta})\ln(1-\hat{\theta})} = \frac{-\overline{X}}{\alpha} \qquad (10)$$

2. Distribution of Households According to the Total Number of Migrants

In this section, some probability models purposed by Singh and Yadava (1991) and Janardan (1973) are applied for describing the variation in households according to total number of out migrants.

Singh and Yadava (1991) proposed a model for the total number of migrants to describe the distribution of households under the following assumptions:

a. Let β be the proportion of households from which at least one person be migrated at the survey point;

b. Out of β proportion of households, let ξ be the proportion of households from which only person be migrated at the survey point;

c. Out of (1- ξ) β proportion of households, let π be the proportion of households from which only males aged 15 years migrate and (1- π) be the proportion of households from which both types of migrants i.e. males above 15 years as well as males with their families migrate;

d. The number of migrants from a household follows a mixture of two displaced geometric distributions with π proportion of households from which only males aged 15 years migrate and (1- π) be the proportion of households from which both type of migration occurs, and

e. Let P_1 be the probability of migration of a person from π proportion of households and P_2 be the probability of migration from (1- π) proportion of households.

Under these assumptions, the probability distribution for the total number of migrants, X (say) is given by

$$p[X=k]=\begin{cases} 1-\beta & \text{if } k=0 \\ \xi\beta & \text{if } k=1 \\ (1-\xi)\beta\left\{\pi p_1 q_1{}^{k-2}+(1-\pi)p_2 q_2{}^{k-2}\right\} & \text{if } k=2,3,4\ldots\ldots\ldots \end{cases} \qquad (11)$$

This model involves five parameters, $\xi, \beta, p_{1,} p_2$ and π to be estimated from the observed distribution of households which is difficult to estimate. In particular, if we assume $p_1 = p_2 = p(\text{say}),$ i.e. risk of migration from the both types of households is same, then model (11) reduces to three parametric model as:

$$p[X=k]=\begin{cases} 1-\beta & \text{if } k=0 \\ \xi\beta & \text{if } k=1 \\ (1-\xi)\beta p q^{k-2} & \text{if } k=2,3,4\ldots\ldots\ldots \end{cases} \qquad (12)$$

The model (12) involves three parameters, ξ, β and p to be estimated.

These are estimated by equating theoretical frequencies to the observed frequencies of first and second cells and theoretical mean to the observed mean i.e.

$$1-\hat{\beta} = \frac{No}{N}$$

$$\hat{\xi}\hat{\beta} = \frac{N_1}{N}$$

and $$\hat{\xi}\hat{\beta} + (1-\hat{\xi})\hat{\beta}\left(1+\frac{1}{\hat{p}}\right) = \bar{x} \quad (13)$$

where $\hat{\xi}, \hat{p}$ and $\hat{\beta}$ denote the estimates of ξ, p and β respectively, No, N1 and N denote the number of observation in zeroth cell, the first cell and sample as a whole respectively and $\bar{x}$ is the observed mean of the distribution.

Janardan (1973) proposed a Generalised Polya Eggenberger Distribution, (GPED) which has a great flexibility for fitting count data. In1999 a new urn model was provided for the GPED. He applied this distribution to a number of situations e.g. the data on the number of defective date fruits, data relating to accidents of 647 women working on H.E shells during a period of five weeks, data relating to lost articles found in the Telephone and Telegraph Buildings, New York, and biological data relating to quadrants of Lespedoza capitata and Liartria aspara, etc. (Green wood and yule (1920)) Generalised Polya Eggenberger Distribution is proposed to give reasonable fit to the data, which follows a negative hypergeometric and also to those data sets for which various modified forms of Poisson, Negative Binomial, Polya of Langragtian- Katz family of distributions have been suggested. Since the occurrence of migration from a household follows a similar situation, it is assumed that this distribution may describe the distribution of the total number of migrants occurring from a household.

Let x be the total number of migrants and follows Polya Eggenberger distribution with probability mass function given by

$$P(x) = \frac{a}{a+bx}(a+bx)^{(x,c)}(1-\beta)^{a+bx)/c},$$

for x =o, 1,2,…., a>0, b>0, b+c>0 and $0<\beta<1$ (14)

Where $m^{(r,c)}$ = m (m+c) (m+2c) ……. (m+xc) and (m^{oc}) =1;

and a, b, c, β are the four parameters of which only three are independent parameters. That is if $\theta = a/c, \eta = b/c$ then the probability mass function is reduced to three parameters model as:

$$P(\gamma,\theta,\eta,\beta) = \frac{\theta}{X!} \quad \{II(\theta+\eta x + j)\}\beta^x (1-\beta)^{(\theta+\eta x)} \tag{15}$$

The first three central moments are used to estimate the three parameters θ, η and β as:

$$\bar{x} = \frac{\theta\beta}{1-(\eta+1)\beta} \tag{16}$$

$$m_2 = \frac{\theta\beta(1-\beta)}{(1-(\eta+1)\beta)^3} \tag{17}$$

$$m_3 = \frac{\theta\beta(1-\beta[1+2\eta\beta-(\eta+1)\beta^2])}{(1-(\eta+1)\beta)^5} \tag{18}$$

$$\text{where } m_1 = \frac{1}{N}\sum_{x=1}^{k}(x-\bar{x})^i \eta_x \tag{19}$$

On simplification, the equations reduces to

$$\beta = \frac{\bar{x}}{\theta+(\eta+1)\bar{x}} \tag{20}$$

$$\frac{m_3}{\bar{x}} = \frac{1-\beta}{(1-(\eta+1)\beta)^2} \tag{21}$$

$$\frac{m_3}{m_2} = \frac{1+2\eta\beta-(\eta+1)\beta^2}{(1-(\eta+1)\beta)^2} \tag{22}$$

Data and Applications

As mentioned above the present study is essentially a comparative study on out migration from rural areas utilizing the data of two surveys conducted in 1978 and

2001. The details of the 1978 and 2001 surveys are given in Sharma (1984) and Shukla (2002) respectively. A brief description of both surveys data is given below.

The data has been collected under a sample survey "Demographic Survey of Chanduali District (Rural area) 2001, India. The survey has been conducted during September 2001 to March 2002. The survey included all the households numbering 402 of selected villages from two blocks, Niyamatabad and Chakia of Chandauli District. At random selection of two villages of Niyamatabad block and four villages of Chakia Bolck of Chanduali District is done.

The 1978 survey was conducted under a research project entitled “Evaluation of the impact of development activities of fertility regulation programme on population growth rate in rural areas”, sponsored by the University Grant Commission, New Delhi, 19 village of Varanasi district which, were classified in three categories as semi-urban, remote and growth center villages representing different levels of development in the rural area.

The inflated geometric and inflated logarithmic distributions are fitted to the data as mentioned above. Table 1 shows the distribution of observed and expected number of households according to the number of migrants male ($\geq$15 years). The value of χ^2 was found not significant at 5% and 1% levels. This suggests that inflated geometric and inflated logarithmic distribution proposed by Sharma (1988) are better approximation to the present situation also.

Table 1: Observed and expected numbers of households according to number of male migrants aged 15 years and above

No. of Migrants	2001 Survey			1978 Survey		
	Observed	Expected		Observed	Expected	
		Inflated Geo.(Ex)	Inflated Logarithmic		Inflated Geo.(Ex)	Inflated Logarithmic
0	242	242.0	242.0	871	871.0	871.0
1	97	97.4	104.5	176	176.0	184.6
2	35	38.5	32.5	59	58.1	48.9
3	19	15.6	14.0	18	19.1	17.3
4	6	8.5	9.0	6	6.3	6.9
5	3			4	3.1	5.3
6	0			0		
7	0			0		
8+	0			0		
Total	**402**	**402.0**	**402.0**	**1134**	**1134.0**	**1134.0**

$\chi^2 = 1.15$, 2.51 0.354 2.945
d.f = 2 2 3 3
α = 0.3980 0.2319
$\hat{p}$ = 0.6084 0.6708
$\hat{\theta}$ = 0.6028 0.53

The models are applied to the data collected during the year 2001 as mentioned above.The model are also compared with the findings obtained by the predecessor using the data of 1978 survey which has been conducted in the same area and under the similar situations. One of the authors of this paper collected data from some developed and undeveloped villages of Chandauli District during 2001-2002.

According to the nature of data surveyed in 2001, it suits for comparison with the remote area's data of 1978 survey as villages from where 1978 and 2001surveys data have been collected are situated at long distances from Varanasi city.

Table 2 shows the distribution of observed and expected number of households by the total number of migrants in 1978 and 2001 surveys. The value of $\chi^2 = 1.85$ was found insignificant at 5% and 1% levels. This shows that the model proposed by Singh and Yadava, 1991, fitted to the distributions of households according to the total number of migrants found in 2001 survey data. An estimate of the proportion of households having only one migrant was found relatively low (0.5125) in 2001 survey in comparison to (0.5285) 1978 survey. This indicates that the chance of migrating singly (leaving their wives and children in the villages) is relatively becoming lower over time.

An estimate of β (proportion of household having at least one migrant) was found more (0.3980) in the 2001 survey than in comparison to (0.2319) found in 1978 survey. This implies that an improved higher status of society has cause a higher rate of migration.

Table 2: Observed and expected numbers of household according to total number of migrants in 1978 and 2001 survey (according to Singh and Yadav, 1991).

No. of Migrants	2001		1978	
	Observed	Expected	Observed	Expected
0	242	242.0	871	871.00
1	82	82.0	139	139.00
2	38	3.0	52	42.00
3	17	20.4	15	27.78
4	11	10.6	14	18.36
5	7	5.8	11	12.14
6	3	3.2	10	8.03
7	2		6	15.85
8	0		16	
Total	**402**	**402**	**1134**	**1134.00**

2001	1978
$\chi^2 = 1.85$	$\chi^2 = 12.17$
$\xi = 0.5125$	$\xi = 0.5285$
$\beta = 0.3980$	$\beta = 0.2319$
$p = 0.4871$	$p = 0.3388$
d.f = 3	d.f =3

Table 3 shows the distribution of observed and expected number of households according to the total number of migrants in both 1978 and 2001 surveys. The value of $\chi^2 = 4.87$ was found insignificant at 5% and 1% levels. Thus the Janardan 1999 model also fitted well to the 2001 survey data, where the value of θ was found higher (1.9067), in comparison to 1978 survey. A higher value of θ indicated a higher chance of being migrated by a male member from a household. Similarly, β indicates the chance of male migrants with their dependants and was found minimum (0.2468) with respect to the value of θ.

The average number of migrants per household was found higher (0.7960) in 2001 survey than those found (0.5547) in the 1978 survey.

Table 3: Observed and expected numbers of household according to total number of migrants in 1978 and 2001 survey (according to Janardan, 1999).

No. of Migrants	2001		1978	
	Observed	Expected	Observed	Expected
0	242	234.11	871	876.26
1	82	91.47	139	147.44
2	38	40.82	52	60.11
3	17	18.93	15	35.42
4	11	7.90	14	11.31
5	7	5.86	11	1.54
6	3	2.80	10	.45
7	2		6	.048
8	0		16	
Total	**402**	**402**	**1134**	**1134**

2001	1978
$\chi^2 = 4.87$	$\chi^2 = 861.65$
$\eta = 0.6555$	$\eta = 0.2045$
$\theta = 1.9067$	$\theta = 0.3691$
$\beta = 0.2468$	$\beta = 0.5347$
d.f = 3	d.f = 4
Average No. of migrants per H.H.	
0.7960	0.5547

Conclusions

A number of models are tested to a new set of data on the number of migrants at the household level. It has been found that models are better approximation to the present situation also. An estimate of proportion of households having at least one migrant has been found maximum in the 2001 survey. It is due to improved higher status of society in rural areas has caused higher rate of migration.

Acknowledgement

We are thankful to the referee for given valuable suggestion to improve the quality of the paper.

References

Green wood, M. and G. Udny Yule (1920) "An inquiry into nature of frequency distribution representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents" *J.Roy,* Ser. A, 83, 255-279.

Hossain, M.Z., 2000, "Some demographic Models and their application with reference to Bangladesh." Unpublished Ph.D. thesis, Department of Statistics, B.H.U., India.

Iwunor, C.C.O., 1995, "Estimating of Parameter of the Inflated Geometric distribution for Rural out Migration" *Genus.* Vol.L1(3-4), 253-260.

Janardan, K. G., 1973, "A four urn Model with Predetermined strategy." Technical report No.73-1, Sangaman State University.

Janardan, K. G., 1999, "Estimation of Parameter of Generalised Polya Eggenberger distribution ." *Commun. Statistics Theory,* Maths, 28(9), 2167-2179.

Kushwaha, S.N.S., 1992, “Some Probability Models for Rural out migration and migrant’s fertility.” Unpublished Ph.D. thesis, Department of Statistics, B. H.U., India.

Sharma, H.L., 1985, “A probability distribution for out migration at micro-level.” *Rural Demography*, Vol.12(1-2), 63-69.

Sharma, H. L., 1988, “Two inflated probability distribution for migration from rural areas.” *Aligarh Journal of Statistics*, Vol.8(1988), 35-41.

Sharma, H.L., 1997, “A Probability Distribution for out migration” *Jansamkhya,* Vol.5(2), 95-101.

Sharma, L., 1984, “A study of the pattern of out migration from rural areas,” an Unpublished Ph.D. thesis, Department of Statistics, B. H. U., India.

Shukla, K. K., 2002, “A study on rural out-migration in Northern India.” Unpublished Ph.D. thesis, Department of Statistics, B. H. U., India.

Singh, A.K. and Kumar, R., 1993, “On the distribution of birth in human population.” *Canadian studies in population*, 20(2).

Singh, P.K., 1992, “Some Demographic Models and their application.” Unpublished Ph.D. thesis, Department of Statistics, B.H.U. India.

Singh S. N. and Yadava; K. N. S., 1991, “Trends in Rural out migration”, *Rural Demography*, Vol.8, No.11, pp.53-61.

Singh, S. R. J., 1985, “A study on Rural migration,” Unpublished Ph.D. thesis, Department of Statistics B. H. U., India.

Singh, S.K., 1990, “On some migration models and their application.” UnpublishedPh.D. thesis, Department of statistics, B. H. U., India.

Yadava, S.N., 1991, “On some Migration and Population Growth Models.” Unpublished Ph.D. thesis, Department of Statistics, B. H. U., India.

Yadava K.N.S., Tripathi, S. and Singh,V.S., 1994, “A Probability Model for the total number of migrants from the Household, An alternative approach.” *Journal of Scientific Research, B. H. U., India* Vol.49, 125-134.

Yadava, K. N. S. and Yadava,G.S., 1980, "On some Probability models and their applications to the distribution of the number of migrants from a household." *Janasamkhya*, Vol.6(2).

Yadava, K. N. S. and Singh.,S.R.J., 1983, "A model for the number of Rural out migrants at households level." *Rural Demography*, Vol.10, 23-33.