Vol. 9 (2005) Journal of International Academy of Physical Sciences pp.47 – 56

Out-Migration : A Regression Analysis

S. R. J. Singh

Department of Mathematics & Statistics, J.N.K. Vishwa Vidyalaya College of Agriculture, Rewa

S. S. Gautam

Department of Statistics, M.G.G. Vishwa Vidyalaya, Chitrakoot, Satna

(Received September 09, 2005)

Abstract. Different models and techniques have been evolved to study the effect of various factors associated with migratory flows from rural areas. In this paper, some regression models have been proposed taking different independent variables which enables exploratory analysis of the significance of numerous factors in the act of migration. Also it provides the means to test the prediction of a strictly specified model of migration. For this study, data are taken from a "Demographic Survey of Rewa-Rural 1998". The present study reveals that the pattern of out-migration from the villages is well followed by multiple linear model proposed on the basis of the prior migrants and other variables in the present situation at the micro level.*

Introduction

Policy makers, planners, social scientists and researchers have focused their increasing attention on the role of out migration as having special significance in the context of rural development. Several models and techniques have been used to study the effect of various factors associated with migratory flows from rural areas. One of these models is multiple regression analysis which enables exploratory analysis of the significance of numerous factors in the process of migration.

It provides the means to test the predictions of a strictly specified model of migration. Its main application has been to investigate the dependency of a measure of migration on one or more explanatory variables, $X_1, X_2, \ldots\ldots X_k$.

Yadava[1] had used a regression analysis to describe the nature of out-migration at village level using variables: (a) number of prior migrants, (b) distance of the village from the nearest large city, (c) percentage of house holds belonging to the upper caste group, (d) the educational level end (e) socio-economic condition of the particular village. These variables explain fifty eight percent of variation in the current flow of migration from villages and the coefficients of the variables of prior number of migrants, distance of the village from the nearest large city and the percentage of households belonging to upper caste group were significant at five percent level whereas the factors, educational level and the socio-economic conditions of the villages were not significant.

As described above, the prior number of migrants is one of the important factors to explain migration (Greenwood[2,3] and Yadava[1]). The researchers have examined the relevance of prior number of migrants with data relating to one point observations. In the present paper first we will examine the role of some other variables in the absence of prior number of migrants. Secondly, we propose to examine the hypothesis, prior migration effects current migration (current migration relates to most recent migration). For this purpose, we have classified the prior number of migrants into two categories having the adequate number of observations in each category. The first and second categories, respectively, consist of recent and remote past prior migrants. The objectives of the present paper is to study :

i) The effect of some variables on current migration in the absence of prior number of migrants,

ii) The effect of recent past prior migrants along with some other variables, on current migration.

iii) The effect of remote past prior migrants along with some other variables, on current migration,

iv) The effect of remote past prior migrants along with some variables on recent past prior migrants, and

v) The validity of the regression model.

The Data

The proposed model is applied to the data collected from rural areas of Rewa district of Madhya Pradesh (M.P.). The data have been collected under a demographic survey of Rewa-Rural 1998 as Dashahara taking reference date. The cluster survey sampling methodology has been adopted to select villages. Out of 2200 sample households from 19 clusters, the survey has covered all the 1789 migrants households within the range of 15 km from the center of Rewa city. A modified definition of household was adopted to cover the information on migration (Singh[4]). The main aim of the survey was to study the existing levels of migration and fertility in the rural areas.

The Regression Model

In this section, some regression models have been proposed taking different independent variables. That is, an attempt is to be made to find out the factors responsible for migration decision process at village level. We have limited the present analysis to eight variables (a) recent past prior number of migrants, (b) remote past prior number of migrants, (c) percentage of those owning land less than four bighas from the particular village, (d) the percentage of households belonging to upper caste group, (e) distance of the village from the nearest town (in Km.), (f) the educational level of the villages, (g) type of family and (h) the diffusion of information.

Incorporating these factors, the regression model can be written in the form

(1) $$Y_i : a_0 + \sum_{j=1}^{8} a_j x_{ji} + E_i \qquad \text{for } i = 1,2,....N$$

where,

Y_i the percentage of out-migrants to the total population from the i^{th} village defined in the period (1994-1998) under consideration,

X_{1i} the percentage of recent past prior migrants to the total population defined in the period (1986-1993) from i^{th} village,

X_{2i} the percentage of remote past prior migrants to the total population defined in the period less than or equal to the year 1985 from i^{th} village,

X_{3i} the percentage of households owning land less than four bighas from i^{th} village,

X_{4i} the percentage of households belonging to upper caste group from i^{th} village,

X_{5i} the distance of the i^{th} village from the nearest town (in km.),

X_{6i} educational level of the i^{th} village,

X_{7i} type of family of the i^{th} village,

X_{8i} the percentage of households which were receiving different kinds of information regarding opportunities at the place of destination from radio, T.V. and news papers in the i^{th} village,

E_i random error having mean zero and a constant variance σ^2.

a_j for j : 0, 1...8, are the parameters of the model to be estimated and N denotes the number of villages.

Measurement of the Variables

Migrants from villages are classified into three groups: The persons who migrated within the period (1994-1998) are termed as current migrants, those migrated within the period (1986-1993) and earlier 1985 respectively, are called recent past prior migrants and remote past prior migrants. The three categories of different migrants are considered keeping in views the small number of observations.

The dependents variable Y_i in the proposed model, is the percentage of current migrants to total population at the time of survey of i^{th} village. On account of the availability of requisite information, independent variables in our study are measured by following Yadava[1]. The numerical values of all the variables and zero order correlation matrix among different variables are given in the tables (1 and 2) respectively.

Table-1: Vales of Y, X_1,X_2,X_3,X_4,X_5,X_6,X_7,X_8 for each village

Village Code	Variables								
	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	0.91	1.17	1.17	21.37	4.96	8.0	0	0	66.94
X_2	1.17	1.80	0.97	33.50	32.41	8.2	1	1	72.22
X_3	1.30	0.91	0.52	48.80	6.60	6.6	0	0	74.02
X_4	3.63	5.01	2.16	37.01	34.06	7.1	1	1	76.09
X_5	3.16	4.36	5.11	45.95	18.94	6.4	1	1	67.39
X_6	1.18	2.16	0.39	55.93	0.00	6.2	1	1	66.67
X_7	1.27	1.39	0.12	25.42	18.90	4.4	0	0	67.72
X_8	1.39	1.39	0.54	37.59	7.39	4.3	0	0	69.72
Y_1	1.58	0.96	0.39	47.92	3.77	11.0	1	1	62.64
Y_2	3.56	2.50	1.10	56.93	7.14	15.8	1	1	72.08
Y_3	4.61	3.90	2.50	59.57	0.00	21.7	1	1	67.86
Y_4	3.78	5.92	3.90	45.98	36.63	16.5	1	1	63.37
Y_5	4.71	3.10	1.33	63.97	11.57	14.8	0	1	48.13
Y_6	4.24	3.48	1.20	21.60	23.18	19.0	1	0	39.07
O_1	5.10	4.40	2.58	43.62	13.90	12.0	1	1	50.66
O_2	1.53	2.07	1.18	23.48	22.70	12.8	0	0	50.66
O_3	1.15	0.42	0.63	36.54	11.52	6.7	0	0	26.67
O_4	2.47	1.82	1.71	54.55	7.33	6.5	0	1	32.00
O_5	1.50	1.41	0.84	30.71	10.61	6.2	0	0	27.37

Table-2: Zero order correlation matrix among different variables

	Variables								
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
X_1	1.000								
X_2	0.817**	1.000							
X_3	0.226	0.221	1.000						
X_4	0.532*	0.379	-0.411	1.000					
X_5	0.506*	0.286	0.253	0.068	1.000				
X_6	0.640**	0.464*	0.254	0.441	0.441	1.000			
X_7	0.674**	0.580**	0.638**	0.175	0.250	0.999**	1.000		
X_8	0.257	0.147	0.154	0.087	-0.014	0.404	0.258	1.000	
Y	0.808**	0.588**	0.417	0.212	0.175**	0.496*	0.597**	0.031	1.000

* Significant at $p = 0.05$, ** Significant at $p = 0.01$

The Regression Findings

After computing the dependent variable Y and independent variables X_j, for J: 1........8, for each village, the corrected sum of squares and product term are calculated. The regression equation for the expectation of percentage of migrants from each village is obtained by least square method. The standard errors of the estimated coefficient are obtained from the variance-covariance matrix. The contribution of the independent variables have been tested by means of the standard t-test. To test the goodness of fit, F-ratios are also computed (see Table 3).

Table-3 : Analysis of Variance Table

Source of variation	*d.f.*	*s.s.*	*m.s.s.*	*F-ratios*	*252552*
$X_3, X_4, X_5, X_6, X_7, X_8$	6	23.7978	3.9663	3.20	$R^2 = 0.62$
Residual	12	14.5052	1.2080 (S^2)		
Total	18	38.3030			
X_1, X_3, X_4, X_5, X_6	5	31.6270	6.3245	12.32	$R^2 = 0.83$
Residual	13	6.6760	0.5135 (S^2)		
Total	18	38.3030			
X_2, X_3, X_4, X_5, X_6	5	26.7914	5.3583	6.05	$R^2 = 0.70$
Residual	13	11.5116	0.8855 (S^2)		
Total	18	38.3030			
X_2, X_3, X_4, X_5, X_6	5	35.2836	7.0567	10.81	$R^2 = 0.81$
Residual	13	8.4864	0.6228 (S^2)		
Total	18	43.77			
X_2, X_3, X_4, X_5, X_6	5	25.3745	8.4582	9.81	$R^2 = 0.66$
Residual	13	12.9285	0.8619 (S^2)		
Total	18	38.3030			

The expected percentage of migrants $\hat{Y}_i$ without considering the prior number of migrants is given by the following regression equation:

$$\text{(2)} \quad \hat{Y}_i : -0.54 + \underset{(.0194)}{0.0419x_{3i}} + \underset{(.0232)}{0.0434x_{4i}} + \underset{(.0499)}{0.1339x_{5i}} + \underset{(.5093)}{0.8325x_{6i}} - \underset{(.5093)}{0.4249x_{7i}} - \underset{(.0160)}{0.0147x_{8i}}$$

From model (2) it is observed that the variables x_7 and x_8 are insignificant and remaining variables are significant at some level. However, the value of R^2 (coefficient of determination) is found to be 0.62 which is satisfactory representation of migration at village level. Dropping the variables whose contributions are insignificant and including one variable i.e. recent past prior number of migrants, a second multiple linear regression model is given below:

$$\text{(3)} \quad \hat{Y}_i : -0.44 + \underset{(.1083)}{0.6889x_{1i}} + \underset{(.0126)}{0.0142x_{3i}} - \underset{(.0151)}{0.0177x_{4i}} + \underset{(.0325)}{0.1096x_{5i}} - \underset{(.3320)}{0.4259x_{6i}}$$

From model (3) the value of R^2 is found to be 0.83 which is more than that obtained by model (2). It is mostly due to recent past prior number of migrants. The coefficient of the variable x_1 and x_5 are significant at 5 percent level. The coefficient of variable x_3 is not so significant.

The third multiple regression model is considered replacing x_1 variable as x_2 in the model (3). It is given below:-

(4) $$\hat{Y}_i : -1.13 + 0.2919x_{2i} + 0.0313x_{3i} + 0.0243x_{4i} + 0.1619x_{5i} - 0.1211x_{6i}$$
$$(.1742) \quad (.0166) \quad (.0198) \quad (.0427) \quad (.4359)$$

From model (4) the value of R^2 is found to be 0.70 which is less than the value obtained from model (3). In this respect, we can say that remote past prior migrants may give less response than recent past. Thus, there is some lagged response due to remote past prior number of migrants than recent past.

In the model (3) and (4), the effect of recent past prior number of migrants and remote past prior migrants respectively are observed on the current migrants. It is interesting to examine the effect of remote past prior migrants on recent past prior number of migrants, and observe the behaviour of this dependency as considered in recent prior number of migrants to current migrants. The regression model is determined and it is as follows:

(5) $$\hat{Y}_i : -0.50+0.5976x_{2i}+0.0131x_{3i}+0.0479x_{4i} + 0.0707x_{5i} + 0,3827x_{6i}$$
$$(.1463) \quad (.0139) \quad (.0166) \quad (.0358) \quad (.3656)$$

The value of R^2 is found to be 0.81 from model (5). It is almost equal to the value of R^2 obtained from model (3).

Since in our analysis the variable x_6 i.e. educational level of the particular village has not played any significant role, we develop the regression model taking only three variables x_3, x_4 and x_5 (it may give the original cause of migration), which is given below:

(6) $$\hat{Y}_i : -1.57+0.0417x_{3i}+0.0472x_{4i}+0.1677x_{5i}$$
$$(.0164) \quad (.0196) \quad (.0421)$$

From model (6) the value of R^2 is found to be 0.66. The coefficients of the variables are significant at 5 percent level. All the models considered here are based on the data of all the migrants from a household.

If we consider the male migrants aged fifteen and above from a household as dependent variable Y, then the model consisting of the variables x_3, x_4 and x_5 becomes:

(7) $$\hat{Y}_i : -7.0568 + 0.1979x_{3i} + 0.2108x_{4i} + 0.8184x_{5i}$$
$$(.0834) \quad (.0999) \quad (.2152)$$

The model (7) also gives almost the same value of R^2 and is equal to 0.64. There is no major difference either we take dependent variable Y as total migrants or male migrants aged fifteen and more from a household.

In this particular situation, the validity of a regression model is also examined. For the regression model, it has been assumed that errors follow independent normal distributions with mean zero and a constant variance s^2.

If the model is appropriate in this situation, then the residual mean square, s^2 (say), is an estimate of variance s^2, and the quantity e_i/s for i: 1,2....N, is called the unit normal deviate from the residual e_i and the quantity e_i/s is expected to lie between -1.96 to +1.96. For each village, the residuals e_i have been computed for models. The expected values $\hat{Y}_i$ of Y_i, residuals e_i and the quantities e_i/s for i: 1,2....N, are given in table 4.

Table-4(a): Residual Analysis for Models 2 and 3

Village Code	Observed Y_i	Model (2) Predicted $\hat{Y}_i$	Model (2) $Y_i - \hat{Y}_i : e_i$	Model (2) e_i/s	Model (3) Predicted $\hat{Y}_i$	Model (3) $Y_i - \hat{Y}_i : e_i$	Model (3) e_i/s
X_1	0.91	0.66	0.25	0.23	1.46	-0.55	-0.77
X_2	1.17	2.71	-1.54	-1.40	1.18	-0.01	-0.01
X_3	1.30	1.57	-20.27	-0.25	1.49	-0.19	-0.27
X_4	3.63	2.73	0.90	0.82	3.29	0.34	0.47
X_5	3.16	2.46	0.70	0.64	3.17	-0.01	-0.01
X_6	1.18	2.06	-0.88	-0.80	1.43	-0.25	-0.35
X_7	1.27	0.94	0.33	0.30	1.03	0.24	0.33
X_8	1.39	0.91	0.48	0.44	1.40	-0.01	-0.01
Y_1	1.58	3.02	-1.44	-1.31	1.61	-0.03	-0.04
Y_2	3.56	3.62	-0.06	-0.05	3.23	0.29	0.40
Y_3	4.16	4.27	0.34	0.31	5.05	-0.44	-0.61
Y_4	3.78	4.66	-0.88	-0.80	5.03	-1.25	-1.74
Y_5	4.71	3.49	1.22	1.11	4.02	0.69	0.96
Y_6	4.24	3.17	1.07	0.97	3.51	0.73	1.02
O_1	5.10	2.98	2.11	1.92	3.85	1.25	1.74
O_2	1.53	2.40	-0.87	-0.79	2.32	-0.79	-1.10
O_3	1.15	2.00	-0.85	-0.77	0.90	0.25	0.35
O_4	2.47	2.04	0.43	0.39	2.17	0.30	0.42
O_5	1.50	1.64	0.14	0.13	1.46	0.04	0.06

Discussion and Conclusions

The analysis of the data in the above sections reveals that the out-migration in the current period is highly related to the recent past prior migration and then remote past prior migration. The simple correlation coefficient between current migration and recent past as well as remote past prior migrants are found to be 0.81 and 0.59 respectively (value of t = 5.70 and t = 3.01 respectively with d.f. 17). As it is apparent from model (3) the value of R^2 is 0.83 which shows the effect of recent past prior migrants along with some other variables on current migration. Similarly, the value of R^2 is 0.70 from model (4) which presents the effect of remote past prior migrants along with some other variables on current migration. These two values of R^2 are quite different which can give the support to the hypothesis that prior migrants affect current migration. It is interesting to determine the effect of remote past prior migrants along with some other variables on recent past prior migrants. In this regard, the value of R^2 from model (5) is found to be 0.81. Thus, the value of R^2 from model (5) is almost equal to the value of R^2 obtained from model (3). However, it is based on the population at the time of survey not on the population of that periods. i.e. the behaviour of future migration depends just upon recent past prior migrants and not on remote past. In this way we can say that prior migrants are the best predicator for migration at micro level. The contribution of recent past prior number of migrants may be adult migrants. They may be eager to know about job opportunities and regularly with new ideas, ways of living etc., which often work as strong incentives for out-migration. So far it is conerned with the validity of the regression models given by equations (3), (4) and (5) the quantity e_i/s is given in table 4. The models (3) and (5) are appropriate while in model (4) one value 2.07 is significant which does not lie between -1.96 to +1.96.

Table-4(b): Residual Analysis for Models 4,5 and 6

Village Code	*Observed* Y_i	*Model (4)*			*Model (5)*			*Model (6)*		
		Predicted $\hat{Y}_i$	$Y_i - \hat{Y}_i : e_i$	$e_i\ s$	*Predicted* $\hat{Y}_i$	$Y_i - \hat{Y}_i : e_i$	$e_i\ s$	*Predicted* $\hat{Y}_i$	$Y_i - \hat{Y}_i : e_i$	$e_i\ s$
X_1	0.91	1.30	-0.39	-0.41	1.28	-0.11	-0.14	0.90	0.01	0.01
X_2	1.17	2.20	-1.03	-1.09	3.03	-1.23	-1.52	2.73	-1.56	-1.78
X_3	3.65	1.77	-0.47	-0.50	1.22	-0.31	-0.38	1.87	-0.57	-0.61
X_4	1.30	3.58	0.05	0.05	3.79	1.22	1.51	2.77	0.86	0.93
X_5	3.16	3.17	0.01	0.01	4.88	-0.52	-0.64	2.29	0.87	0.94
X_6	1.18	1.62	-0.44	-0.47	1.29	0.87	1.08	1.80	-0.62	-0.67
X_7	1.27	0.87	0.40	0.43	1.12	0.27	0.33	1.12	0.15	0.16
X_8	1.39	1.08	0.31	0.33	0.97	0.42	0.52	1.07	0.32	0.34
Y_1	1.58	2.24	-0.66	-0.70	14.70	-0.74	-0.92	2.45	-0.87	-0.94
Y_2	3.56	3.60	-0.04	-0.04	2.78	-0.28	-0.35	3.79	-0.23	-0.25
Y_3	4.61	4.86	-0.25	-0.27	3.69	0.21	0.26	4.55	0.06	0.06
Y_4	3.78	4.89	-1.11	-1.18	5.74	0.18	0.22	4.84	-1.06	-1.14
Y_5	4.71	3.93	0.77	0.82	2.73	0.37	0.46	4.13	0.58	0.62
Y_6	4.24	3.41	0.83	0.88	3.34	0.14	0.17	3.61	0.63	0.68
O_1	5.10	3.15	1.95	2.07*	4.36	0.04	0.05	2.92	2.18	2.35**
O_2	1.53	2.57	-1.04	-1.11	2.50	0.43	0.53	2.63	-1.10	-1.18
O_3	1.15	1.56	-0.41	-0.44	1.38	-0.96	-1.19	1.62	-0.47	-0.51
O_4	2.47	2.31	0.16	0.17	2.05	-0.23	-0.28	2.17	0.30	0.32
O_5	1.50	1.34	0.16	0.17	1.35	0.06	0.07	1.25	0.25	0.27

The value of R^2 from model (2) is found to be 0.62 consisting of x_3, x_4, x_5, x_6, x_7 and x_8 variables, while it is 0.66 consisting of only three variables x_3, x_4 and x_5. This may be due to the fact that the variables x_6 and x_7 are measured by dummy variables as 0 and 1 which is a crude measure. Also, it arises due to multicollinearity among some variables. However, the difference is quite small in the values of both R^2. So far it is concerned with the validty of the regression models, the model (2) is appropriate while in model (6) one value 2.35 is significant which does not lie between -1.96 to +1.96.

The third vaiable which is the percentage owning land less than four bighas from the particular village is positively correlated with current migration. It is found to be significant at 5% level in model (7).

The fourth variable is the percentage of household belonging to the upper caste group. The joint family system is relatively very common in these castes. If there is a large number of members, in the working group of households, they are more prone to migration. Therefore migration is relatively more common in these castes.

The fifth significant variable is the distance of a village from the nearest town. The villages near the Rewa City reduce the out-migrants while the villages far from the city of Rewa induce the in-migrants (the simple correlation between distance and current migration is 0.66, t: 5.49 with 17 d.f.). This indicates that this variable is highly related to the current migration.

The coefficient of the next variable, i.e., educational level of the village is not so significant in the models. This may be due to the drawback in way of measuring this variable.

As it is mentioned earlier that the variables of x_6 and x_7 are found to be insignificant. In the case of multicollinearity, probably the best that could be done is to drop the insignificant variables. Also these variables are measured by dummy variables as 0 and 1, which is the crude measure. The findings do not mean that the relationships between Y and x_6 or x_7 are non-existent in the population. They suggest the need to improve the process of measurement of these variables.

However, it is clear from the present study that the pattern of out-migration from the villages is well followed my multiple linear regression model proposed on the basis of the prior migrants and other variables in the present situation at the micro level.

References

1. K.N.S. Yadava: Some aspects of out-migration and its impact on fertility, Unpublished Ph.D. thesis in statistics, Banaras Hindu University, Varanasi , 1977.
2. M.J. Greenwood: The determinants of labour migration in Egypt, *Journal of Regional Sciences*, 9 (1969) 283.

3. M.J. Greenwood: Lagged response in the division to migrate, *Journal of Regional Sciences*, **10** (1970) 375.

4. S.R.J. Singh: A study of rural out migration and its effect on fertility, Unpublished Ph.D. thesis in statistics, Banaras Hindu University, Varanasi, 1985.